TOWARDS CONSISTENT TRAVEL DEMAND ESTIMATION IN TRANSPORTATION PLANNING

A GUIDE TO THE THEORY AND PRACTICE OF EQUILIBRIUM TRAVEL DEMAND MODELING

FINAL REPORT

June 27, 2001

U.S. Department of Transportation

Bureau of Transportation Statistics

Prepared by

Harvey J. Miller
Department of Geography
University of Utah
Salt Lake City, Utah 84112
harvey.miller@geog.utah.edu

1. EXECUTIVE SUMMARY

A travel demand analysis consists of assessing four components of the travel pattern for a study area: i) *trip generation* (TG) or where trips are coming from; ii) *trip distribution* (TD) or where trips are going to; iii) *modal split* (MS) or the shares among available modes for the flow between origin-destination (O-D) pairs; and, iv) *network assignment* (NA) or the route choice within each mode. The standard "state-of-the-practice" in travel demand modeling is the sequential or *four-step approach*. This modeling strategy estimates the four travel demand components sequentially and feeds the results from one component to the next component in the sequence. Unfortunately, the four-step approach is flawed. A particularly severe problem is potential inconsistency among the travel demand component estimates. Another problem is that prediction errors from any component are compounded in each subsequent stage, potentially leading to substantial errors in latter stages.

While the inherent flaws in the four-step approach are recognized widely, an existing, viable alternative is not widely known. The *equilibrium travel demand modeling* approach embeds the travel demand components in the four-step approach within a *market equilibrium* framework. This generates consistent answers among the four travel demand components. The familiar components from the four-stage approach are preserved; the additional theory and modeling framework simply enforces consistency among these components. *Achieving this consistency in general does not require substantial increases in computational resources nor data inputs.*

This research report addresses the gap between the state-of-the-art and the state-of-the-practice in travel demand modeling. This report is an *accessible* review of the theory and practice of equilibrium travel demand modeling. This review is intended for practitioners and beginning students in transportation analysis, modeling and planning. Key features of this review include: i) a focus on *practical* travel demand models, i.e., models that can be implemented at the urban or regional-scale; ii) a focus on the behavioral assumptions, data requirements, parameter estimation procedures and solution procedures that are key to model application.; iii) placement of mathematical formulae are in appendices, allowing the less mathematically-incline reader to skip the formulae but still receive an intuitive understanding of the models' structures. These features should render this review accessible to its intended audience, transportation analysts and planners.

This report first reviews the theoretical conditions for network flow equilibrium (i.e., the NA phase of the four-step approach) and the overall market equilibrium for the remaining travel demand components (MS, TD, TG) based on the assumed network equilibrium. The available network equilibrium principles include:

i) *User optimal-strict* (UO-S): At network equilibrium, no traveler can reduce his or her travel costs by unilaterally changing routes (i.e., changing routes independently without other users' route changes);

ii) *User optimal-general* (UO-G): Travelers change routes *in the next time* period in a manner that reduces total cost *based on the current route costs*;

iii) *Dynamic user optimal* (DUO): At network equilibrium, no traveler *who departed during the same time interval* can reduce his or her travel costs by unilaterally changing routes;

iv) *Stochastic user optimal* (SUO): At network equilibrium, no traveler can reduce his or her *perceived* travel costs by unilaterally changing routes.

These network equilibrium principles can be linked in a theoretically consistent manner to equilibrium conditions for the higher-level travel demands. Models that do not enforce this simultaneous equilibrium are misspecified and consequently flawed. Empirical evidence going as far back as the 1970's suggests that the four-step approach suffers from misspecification, nonconvergence and error.

Combined travel demand models can be derived based on the each assumed network equilibrium. The following table summarizes the models reviewed in this report:

	Travel demand components			
Network equilibrium class	NA	NA/MS	NA/MS/TD	NA/MS/TD/TG
UO-S	Sheffi (1985)	Evans (1976)	Florian and Nguyen (1978)	STEM (Safwat and Magnanti 1988)
UO-G	T2 (Dial 1995b)	Dafermos (1980)		Dafermos (1982)
DUO	Janson (1991 a,b)			
SUO	Fisk (1980)	Super- and hyper-networks (Sheffi and Daganzo 1980)		
UO-S/SUO	Trip consumer approach (Oppenheim 1995)			

The equilibrium travel demand models discussed generally follow an equivalent optimization approach. This strategy first specifies a combined travel demand model then derives an equivalent optimization problem whose solution corresponds to a market equilibrium of the specified travel demand components in the initial model. Typically, this problem contains a objective function to be minimized and constraints that represent flow and aggregate demand feasibility requirements. In the interest of brevity and due to the pragmatic orientation of this report, this section only discusses the equivalent optimization problems. The report discusses the models from the perspective of: i) *basic assumptions*; ii) *model structure*; iii) *data requirements and parameter estimation*; and iv) *solution procedure*.

After discussing the basic characteristics of each model, this report compares the travel demand methods based on several criteria. The comparison provides guidance for model selection and use, although it does not provide a definitive answer. Comparison of the travel demand models uses the following criteria: i) *basic theory* or the major strengths and weaknesses of the model's theoretical base; ii) *mathematical elegance* or the parsimony and flexibility of the model's formalism; iii) *computational requirements and performance,* including the basic procedural needs of each model's algorithm as well as performance efficiency; and, iv) *data requirements and parameter estimation*.

Although this report’s objective is an accessible review of equilibrium travel demand models rather than the research frontiers, this review nevertheless suggests three major research and development issues. This includes: i) *specification and development of a computational toolkit for equilibrium travel demand modeling*; ii) *development of a travel demand model testbed;* and, iii*) development of a combined statistical distribution theory and simultaneous parameter estimation procedures*. The first issue concerns the specification and development of a toolkit that can support *several* of the equilibrium travel demand models within the same computational platform. Instead of forcing a travel demand analysis into the model available within a given GIS software, this would allow the practitioner to access the model or models most appropriate for the research question at hand. The second issue, closely related to the first, concerns support for extensive testing of equilibrium travel demand models as well as other competing approaches.

The third research and development issue addresses a weakness of equilibrium travel demand models, specifically, a lack of statistical distribution theory for the combined travel demand components within each equilibrium model. This weakness is shared with the 4-step approach: a consistent combined statistical distribution theory does not exist for the sequential travel demand estimation procedure. However, this weakness is not as apparent in the 4-step approach since it artificially separates the travel demand modeling components. When these components are embedded in an equilibrium framework, this weakness becomes more obvious. Some discussion of these combined estimation issues does exist in the literature. However, no existing model has a combined statistical distribution theory and an efficient and unbiased simultaneous estimation procedure for all parameters. Continued research along this line is required for effective application of equilibrium travel demand models.

2. ACKNOWLEDGEMENTS

The Bureau of Transportation Statistics supported this project through a research fellowship and residency at the BTS office in Washington, DC. Thanks to Dr. T.R. Lakshmanan (Director) and Dr. Bruce Spear (Assistant Director for Geographic Information Systems) for facilitating the research fellowship. Dr. Bruce Spear and Dr. Frank Southworth (Oak Ridge National Laboratory) carefully read an earlier version of this manuscript and provided helpful suggestions and comments. All remaining omissions and misstatements are the author's responsibility.

3. TABLE OF CONTENTS

4. LIST OF TABLES

5. LIST OF FIGURES

6. INTRODUCTION

A travel demand analysis consists of assessing four components of the travel pattern for a study area: i) *trip generation* (TG) or where trips are coming from; ii) *trip distribution* (TD) or where trips are going to; iii) *modal split* (MS) or the shares among available modes for the flow between origin-destination (O-D) pairs; and, iv) *network assignment* (NA) or the route choice within each mode. Assessing these components provides insight into the effectiveness of transportation policy and the performance of transportation infrastructure. In addition, being able to *predict* these components through modeling can allow the planner or policy analyst to pose "what-if?" scenarios regarding infrastructure, land use and/or policy changes and estimate the resulting impacts on travel patterns.

The standard "state-of-the-practice" in travel demand modeling is the sequential or *four-step approach*. This modeling strategy estimates the four travel demand components sequentially and feeds the results from one component to the next component in the sequence. A common sequence is TG⧗TD⧗MS⧗NA; this approach is embodied in the *urban transportation modeling system* (UTMS) strategy and commercial software that implement this strategy.

Unfortunately, the four-step approach is flawed. A particularly severe problem is potential inconsistency among the travel demand component estimates. Since the four-step approach does not require internal consistency among the four estimated components, it is not likely to occur in practice. For example, the NA phase changes network travel costs which will no longer be consistent with the travel costs used for the TG, TD and MS phases. A common response is to use feedback loops and "cycle-back" answers to previous steps for additional rounds of estimation. However, this still does not guarantee convergence to a consistent answer. Another problem is that prediction errors from any component are compounded in each subsequent stage, potentially leading to substantial errors in latter stages (Fernandez and Friesz 1983; Sheppard 1995). These theoretical flaws have been substantiated by empirical evidence. As far back as the mid-seventies, Florian, Nguyen and Ferland (1975) found empirical evidence that sequential estimation with feedback of TD-NA does not converge. More recently, empirical benchmarking by COMSIS (1996) and Boyce, Zhang and Lupa (1994) found that sequential estimation results in inferior predictions of key output variables such as traffic flow on links.

The inherent weakness of the four-step approach is recognized widely. For example, the U.S. Department of Transportation, in cooperation with the U.S. Environmental Protection Agency and U.S. Department of Energy, has created the *travel model improvement program* (TMIP) to respond to requirements of the 1991 Clean Air Act and the 1991 Intermodal Surface Transportation Efficiency (ISTEA) Act. A short-term improvement identified by this program is improving the feedback loop strategy in the four-step approach. Long-term improvements include the development of the "next-generation" of travel demand models, i.e., moving beyond the four-step approach. One initiative from the long-term improvement program is TRANSIMS, a cellular automata-based microsimulation travel model (Barrett *et al.* 1995).

While the inherent flaws in the four-step approach are recognized widely, alternatives to this approach are not widely known. Although next-generation modeling initiatives have great merit, there is little awareness of an existing modeling strategy that directly addresses the major flaws in the four-step approach. The modeling strategy is the *equilibrium travel demand modeling* approach. This strategy embeds the travel demand components in the four-step approach within a *market equilibrium* framework. This "simultaneous estimation" approach generates consistent answers among the four travel demand components. The familiar components from the four-stage approach are preserved; the additional theory and modeling framework simply enforces consistency among these components. *Achieving this consistency in general does not require substantial increases in computational resources nor data inputs.* Thus, planners and analyst who are familiar with the four-stage approach can easily understand the equilibrium approach.

The equilibrium travel demand approach has been present in the literature for over three decades. The initial theory was developed in the fifties (Beckmann, McGuire and Winsten 1956; Wardrop 1952). Practical models have existed since the mid-seventies (Evans 1976; Florian and Nguyen 1978). Recently, substantial improvements in this approach have been achieved; these improvements include: i) encompassing all four travel demand components (Safwat and Magnanti 1988); ii) linking the travel demand equilibrium to individual-level choice theory (Oppenheim 1995); and, iii) improving the network flow principles at the basis of the market equilibrium, including more realistic treatment of route choice behavior (Dial 1995a, 1995b, 1996) and extensions to dynamic network flows (Janson 1991a, 1991b).

Despite the long history of equilibrium travel demand models and the flurry of recent research progress, this practical modeling strategy is almost completely unknown to practitioners and not widely known even to transportation academics. Part of the reason is undoubtedly due to inertia created through the UTMS initiative and readily available software that implement this strategy. However, a very large part of the blame must rest with the academics and scientists who develop the state-of-the-art in these models but do not attempt to disseminate this information to practitioners. While several *excellent* reviews of the equilibrium approach exist (Boyce 1984; Boyce, LeBlanc and Chon 1988; Fernandez and Friesz 1983; Friesz 1985) these are somewhat dated and (more importantly) are oriented towards academics and scientists who are interested in the extending the modeling frontier. Consequently, these reviews are not accessible to practitioners. This lack of information flow has hampered the improvement of the state-of-the-practice in travel demand modeling.

This research report attempts to address the gap between the state-of-the-art and the state-of-the-practice in travel demand modeling. This report is an *accessible* review of the theory and practice of equilibrium travel demand modeling. This review is intended for practitioners and beginning students in transportation analysis, modeling and planning. Key features of this review include:

i) A focus on *practical* travel demand models. In this case, "practical" refers to models that can be implemented at the urban or regional-scale without undue computational or data requirements beyond the four-step approach. This selective review focuses on these practical models, although in two cases currently impractical but very promising models are included for completeness.

ii) Discussion focuses on the behavioral assumptions, data requirements, parameter estimation procedures and solution procedures that are key to model application.

iii) Mathematical formulae are placed in appendices and are cross-referenced and explained verbally within the report body. This allows the less mathematically-inclined reader to skip the formulae but still receive an intuitive understanding of the models' structures. Conversely, the more mathematically-inclined reader can follow the cross-references to the corresponding formulae. The cross-referencing system uses an equation labeling scheme that maintains the section and equation number (e.g., (12-1) is Section 12, equation 1).

These features should render this review accessible to its intended audience, transportation analysts and planners.

Section 7 of this report explains the basic theory underlying equilibrium travel demand modeling. This includes discussion of: i) basic transportation system elements; ii) different types of network equilibria; and iii) detailed discussion of travel demand market equilibrium and the weaknesses of the four-step approach. Section 8 constitutes the major portion of this report. This section discusses several major equilibrium travel demand models, with the discussion organized by the type of network equilibrium at each model's basis. Discussion of each model includes; i) major assumptions; ii) model structure; iii) data requirements and parameter estimation; and, iv) solution procedure. Section 9 provides a summary of the models reviewed. This includes a discussion of the major strengths and weaknesses of each model from the perspective of: i) basic theory; ii) mathematical elegance; iii) computational performance; and, iv) data and parameter estimation. Although the research frontier is not the focus of this report, Section 9 also provides some comments on continued research and development needs that can facilitate the more widespread usage of these models. Section 10 provides some brief concluding comments.

7. BASIC THEORY

7.1 Transportation Systems as Markets

The basic idea underlying the network equilibrium approach to travel demand modeling is a view of transportation systems as *markets*. The network equilibrium approach embeds the elements typically found in the traditional, four-step approach into a *market equilibrium framework*. As in classical economic market theory, the task is to predict the short-run equilibrium levels of supply and demand, that is, the number of trips and level of transportation service in the study area (Fernandez and Friesz 1983).

While the concept of market equilibrium is straightforward, its application to transportation systems involves special considerations related to two features of these systems: i) the network basis of transportation systems; and, ii) the existence of demand externalities in the form of *congestion*. In the first case, supply functions are tied to network links; for example, consider the performance functions typically used to relate flow in a link to its travel time or cost (see below). However, the relevant unit of analysis is the individual trip between an origin and a destination; this trip will use a path consisting of multiple network links. Therefore, the equilibrium model must relate each transportation demand (i.e., trip) to multiple supply components. In the second case, each traveler's choice relates to the level of service provided by available paths between an O-D pair. In turn, these service levels are influenced by the choices of other travelers since the performance of service typically degrades as the number of users increase. These congestion externalities suggest that the level of service (supply) and flow (demand) between an origin-destination pair must, *in general*, consider the service levels and flows for all origin-destination pairs in the network (Fernandez and Friesz 1983).

Transportation market equilibrium occurs at two levels. First, the flows through the network correspond to some stated equilibrium criterion such as Wardrop's user-optimal principle (informally, no user can improve his/her cost by unilaterally changing routes; see below). This pattern corresponds to the NA phase of the traditional four step approach. The second equilibrium level corresponds to the TG, TD and MS components of the four step approach. Demand for these "higher-level" components is *elastic*, meaning that it is responsive to cost. Therefore, we can also specify a corresponding market equilibrium criterion at this level, e.g., no user can improve his/her travel cost by unilaterally changing generation rate, destination choice or mode choice. Note that these are tightly linked with the "lower level" network equilibrium since network flow costs affect the higher level demands while the higher level demands affect the amount of network congestion and therefore the network travel costs.

7.2 Basic Components

7.2.1 *Network Characteristics*

A *directed graph* represents the transportation system in the study area. The directed graph consists of a set of network nodes and a set of directed (i.e., "one-way") arcs connecting certain nodes (12-1). Some nodes represent travel origins (12-2) while others represent travel destinations (12-3); the remaining generally correspond to street intersections, modal transfer points and other flow "transfer" locations. A two node sequence represents each network arc in the standard "from-node, to-node" format (12-4).

Sequences of network arcs comprise network paths. These paths originate at origin nodes, terminate at destination nodes and are connected in the sense that the "to- node" of an arc is the "from-node" of the next arc in the sequence (12-5). An *arc-path incidence variable* indicates the relationship between individual arcs and paths (12-9): models use this variable directly to maintain consistent relationships between flows at the arc and path levels.

Travel demand models differ with respect to representation of multiple modes. Often, a model will use a single directed graph to represent all modal networks in a study area. In this case, different modal flows coexist within the same arc (12-10) or within the same path (12-15). In other cases, an explicit multimodal network is required, that is, each mode has a separate directed graph. "Transfer arcs" link these distinct modal networks.

Travel demand models estimate flows at either the arc or path level. *Flow feasibility requirements* ensure that solutions are realistic, consistent between the arc and path levels, and consistent with aggregate-level travel demands (that is, the known or estimated aggregate flows between O-D pairs). These requirements are: i) all path flows are non-negative (12-56); ii) the mode-specific flows on all paths between an O-D pair sum to the aggregate modal flows between that pair (12-57); and, iii) the mode-specific flows on all paths that use an arc sum to the total modal flow on that arc (12-58).

Although travel demand models require consistency between flows at the arc and path level, an interesting theoretical result is that at equilibrium only arc flows and aggregate travel demands are unique: path flows are *not* unique (see Fernandez and Friesz 1983; Sheffi 1985, 66-69). That is, any set of path flows that are consistent with the equilibrium arc flows is allowable; in theory, this is an infinite set. From a practical perspective, this is not a major problem since we are primarily concerned with flow levels within given elements of the transportation infrastructure. However, one must keep in mind that path flow estimates from these models are not suitable for analysis.

7.2.2 *Cost Functions*

Similar to network flows, travel costs can be measured at the arc (12-13) or path levels. However, cost functions are usually specified at the arc level: path travel costs are simply the summed costs for all arcs that comprise that path (12-17).

Mode-specific arc travel costs are generally a function of flow, either the mode-specific flow on that arc (12-59) or a function of all modal flows across all arcs in the network (12-60). The former cost function is referred to as *separable*, i.e., the flows across different modes and different arcs can be meaningfully separated. The latter cost function is referred to as *non-separable*, i.e., flows across different modes and different arcs cannot be partitioned meaningfully into independent flows. Separable cost functions are not as realistic as non-separable functions. For example, a separable cost function assumes that different modes sharing an arc do not influence each other (e.g., automobile congestion on a link does not influence buses using the same link). Also, separable cost functions do not account for the interactions of flows on different arcs (e.g., congestion at intersections due to cross-traffic, interactions among two-way flows on a street). Non-separable cost functions can consider these interactions; however, solving the resulting travel demand model is much more difficult.

Typically, arc flow costs functions represent the *generalized cost* of travel within that element of the transportation infrastructure. For separable cost functions, a basic but typically invoked function is:

$$c_a^k\left(f_a^k\right) = d_a^k + w\ s_a^k\left(f_a^k\right) \qquad \textbf{(7-1)}$$

where d_a^k is the out-of-pocket expense required for using mode k on arc a (this may also be a function of flow), $s_a^k\left(f_a^k\right)$ is the mode k travel time on arc a associated with flow level f_a^k, the mode k flow on arc a, and w is a *value-of-time* (VOT) parameter that translates travel time into equivalent monetary units, i.e., travelers' time cost. One of the two elements of (7-1) may not be present for a given mode within a given arc. For example, public transit fares may only be invoked in modal entry or transfer arcs. The VOT parameter may also be associated with the monetary expense variable instead of the travel time variable, if desired.

Various models require different restrictions on the behavior of cost functions with respect to flow levels. Given the basic format in (7-1), these restrictions are through the flow-based travel time function $s_a^k\left(f_a^k\right)$. A typically invoked format for this function is (Branston 1976):

$$s_a^k\left(f_a^k\right) = \bar{s}_a^k\left(1 + b_1\left(\frac{f_a^k}{B_a^k}\right)^{b_2}\right) \qquad \textbf{(7-2)}$$

where $\bar{s}_a^k$ is the free-flow travel time, B_a^k is the mode k capacity of arc a, and b_1 ,b_2 are empirically-estimated parameters.

7.2.3 *Demand Functions*

Demand functions relate the amount of O-D flow for each mode to travel costs. As with the arc cost functions, demand functions are either separable or non-separable. Separable demand functions relate the level of mode-specific flow between an O-D pair to the minimum cost for that mode and O-D pair only (12-61). In contrast, non-separable demand functions relate the mode-specific flow between an O-D pair to the minimum travel costs across all O-D pairs and modes (12-62). As with arc cost functions, non-separable demand functions are more realistic but result in model formulations that are more difficult to solve.

In some models, the O-D demands are fixed and exogenous, meaning that aggregate O-D flows are required as external data rather than predicted as a model outcome. Evans (1976) provides an example of an endogenous, separable demand function:

$$D_{ij} = A_i \; B_j \; \exp\left(-g_{\,ij} \; C_{ij*}\right) \tag{7-3}$$

where D_{ij} is the aggregate flow between origin i and destination j, C_{ij*} is the minimum travel cost between the O-D pair, g_{ij} is an estimated parameter, and A_i and B_j are "balancing factors" or parameters chosen so that outflows from origins and inflows to destinations sum to totals known from exogenous data (i.e., the total amount of travelers leaving each origin and the total amount of travelers entering each destination). This demand function is essentially a doubly constrained spatial interaction model, that is, a spatial interaction models whose origin outflows and destination inflows are constrained to match known sums (see Fotheringham and O'Kelly 1989; Wilson 1967, 1974).

7.3 Types of Transportation Equilibria

7.3.1 *Network equilibria*

7.3.1.1 User optimal (UO)

7.3.1.1.1 User optimal-strict (UO-S)

The most common type of network equilibria analyzed is the *user optimal* (UO), originally due the Wardrop (1952). The traditional, strict definition, referred hereafter as *user optimal-strict* (UO-S), is:

> (UO-S) At network equilibrium, no traveler can reduce his or her travel costs by unilaterally changing routes (i.e., independently change routes without other users' route changes).
>
> Alternatively: All used routes between an O-D pair have the same, minimal cost and no unused route has a lower cost.

This implies the following network flow characteristics. First, positive flow for a mode on a route implies that it must have a travel cost equal to the minimum cost for that mode between the particular O-D (13-1). Second, any route with a cost greater than the minimum for a mode implies that the flow level for that mode is zero on that route (13-2). In other words, for each mode, flow only occurs on the minimum cost routes between each O-D pair, i.e., no traveler has a less costly alternative route (Smith 1979).

The UO-S conditions imply a tenable behavioral motivation but require strong assumptions about travelers *reactions* to conditions within the network. The fundamental behavioral postulate is that travelers follow the "cheapest" available route for their user class. While this basic motivation seems reasonable, strict adherence to this behavior at an individual-level is less tenable. The UO-S conditions imply travelers' perfect decision-making capabilities and perfect knowledge about network conditions. In other words, travelers know the exact cost on each available route and react to these costs with perfect accuracy. Nevertheless, many equilibrium models follow the UO-S conditions since they result in tractable formulations and provide a "best-guess" about traveler decisions lacking other behavioral data.

7.3.1.1.2 UO-General (UO-G)

Smith (1979) proposed a generalization of the UO conditions that imply less strict behavioral assumptions. Paraphrasing slightly, the *user equilibrium-general* (UO-G) conditions are:

> (UO-G) Travelers change routes *in the next time* period in a manner that reduces total cost *based on the current route costs.*

Travelers change routes in the next time period (e.g., "tomorrow") based on the current time period's costs (e.g., "today"). Therefore, travelers do not react to network conditions instantaneously. Also, the current time period flow pattern influences, but does not determine, the flow pattern in the next time period. In general, a number of flow patterns rather than a single flow pattern in the next period will satisfy UO-G (Fernandez and Friesz 1983; Smith 1979).

Despite the temporal element in the definition, this principle can also characterize static flow patterns since it describes conditions for flow stability. The UO-G conditions state that a flow pattern is UO if any other flow pattern would result in higher total costs (13-3). This expands the UO-S conditions. If the network is at UO-S, the flow pattern will be stable since no traveler can switch routes in the next time period and reduce total cost. However, UO-G also allows flow patterns that do not satisfy UO-S but nevertheless are reasonable from a behavioral perspective. Under UO-G, individual travelers switch to more expensive routes only if that change does not lead to an increase in total cost across all travelers. Thus, some travelers are allowed to make "mistakes" if this does not "harm" other travelers *in toto.*

7.3.1.2 Dynamic User Optimal (DUO)

Static equilibria assume that the travel demand pattern in a given study area converge to a "steady-state" condition in which temporal fluctuations do not occur. Analysts recognize that temporal fluctuations in NA, MS, TD and TG do occur in reality. Since transportation planning is oriented traditionally towards infrastructure planning and broad policy evaluation, ignoring minor temporal fluctuations is defensible since these plans and policies attempt to accommodate the general travel demand pattern.

There have been recent attempts to incorporate dynamic properties of travel demand patterns. These attempts are motivated by the U.S. federal policy shifts away from large infrastructure investments in urban area. Manifestations of this policy shift such as *intelligent transportation systems* (ITS) require detailed temporal predictions of traffic flows and congestion and the implementation of non-transportation, activity-based solutions such as flex-time and telecommuting. Another motivation for dynamic travel demand models is the difficulty in capturing adequately the environmental impacts of travel demand patterns in general and traffic congestion specifically. Assessing the impacts of traffic on ambient air quality requires estimates of behaviors such as engine cold-starts and speed variations. These factors affect air quality more than aggregate throughput *per se* (Kulkarni *et al.* 1996).

"Equilibrium" is a much broader concept in the dynamic realm. A fundamental consideration is the equilibrium's time frame. Time can be viewed as *discrete* (i.e., divided into finite intervals) or *continuous*. In addition, equilibrium conditions can be stated for "within-day" or *intra-periodic*, "day-to-day" or *inter-periodic* or *combined intra/inter-periodic* dynamics. Within-day dynamics capture daily fluctuations in travel demand both with respect to inherent fluctuations as well as unplanned disturbances such as road closings, accidents, etc. Within-day dynamics also allows modeling timing decisions for trip generation; this is important for discretionary travel as well as flex-time-based commuting in congested networks. Day-to-day dynamics capture the slower learning process of travelers as they acquire information about the travel environment. In addition, the existence of a traditional transportation equilibrium is not guaranteed, particularly with respect to continuous time dynamics. The system may converge to different attractors and display complex behavior as with dynamical systems in general (Cantrella and Cascetta 1995).

As the discussion in the previous paragraph implies, there is a wide-range of dynamic equilibrium formulations (e.g., Cantrella and Cascetta 1995; Friesz *et al.* 1994; Friesz, Bernstein and Stough 1996; Ran and Boyce 1994; Ran, Hall and Boyce 1996). Several of the continuous time formulations have similar structure to the UO-G conditions (more specifically, they share the structure of a *variational inequality* problem; see Nagurney 1993). Many are oriented specifically towards ITS rather than travel demand prediction. For example, the formulations of Ran and Boyce (1994) and Ran, Hall and Boyce (1996) assume that the amount of flow entering each transportation link are control variables in their dynamical system. Real-world manifestations of these variables could be traffic control and ITS devices such as variable message signs, variable time traffic lights and information provided to drivers that influence or direct their route choices.

Due to the orientation of this review towards pragmatic travel demand models, the dynamic user optimal (DUO) principle considered here is a discrete-time, within-day formulation by Janson (1991a, 1991b). This DUO formulation has resulted in a very practical dynamic NA model and solution procedure (to be discussed later). This DUO principle is:

> (DUO) At network equilibrium, no traveler *who departed during the same time interval* can reduce his or her travel costs by unilaterally changing routes.
>
> Alternatively: All used routes between an O-D pair have the same, minimal cost and no unused route has a lower cost *for travelers that departed during the same time interval*.

This DUO principle implies the following network flow characteristics. First, positive flow on a route for users who departed during a given time interval implies that it must have a travel cost equal to the minimum cost for those users between the particular O-D pair (13-4). Second, any route with a cost greater than the minimum for users who departed during a given time interval implies that the flow level for those users is zero (13-5). Note that these conditions are a direct extension of the UO-S conditions. Indeed, UO-S is a special case of this DUO principle (Janson 1991).

DUO implies the same harsh behavioral assumptions as UO-S. In addition, the treatment of time as discrete limits the resolution of these dynamics. However, the introduction of a dynamic component increases the realism and usefulness of the UO equilibrium principle. Also, as stated above, this equilibrium principle does allow for a pragmatic dynamic NA model that is tractable computationally and has reasonable data requirements.

7.3.1.3 System Optimal (SO)

While the UO conditions minimize individual travel costs, it does not in general minimize *total* cost for travelers as a whole. The UO-S conditions only require that the flow pattern minimizes costs on an individual basis. The UO-G conditions allow flow changes that do not increase total cost but do not require this to be minimal for individuals. Minimizing individual costs does not equate to minimizing total costs when congestion is present in the network. Under these conditions, each traveler's route choice influences the costs of other travelers.

The UO-S principle assumes that travelers' do not consider the externalities of their decisions: travelers only perceive their personal travel cost and not the additional costs imposed on others by their route choices (Ortuzar and Willumsen 1990). To accommodate this additional decision principle, Wardrop (1952) formulated a second, *system optimal* (SO) principle:

(SO) At network equilibrium, the total (or average) travel cost is minimum.

A flow pattern that satisfies this principal is appealing from a society-wide perspective. An SO flow minimizes the total operating cost of the network, implying efficiency (Fernandez and Friesz 1983). Also, if we accept total cost as a surrogate for the system-wide use of energy resources and output of pollution, we can see that this pattern would minimize these negative impacts. However, this flow pattern is not likely to occur in practice since it requires travelers to make joint decisions to minimize total cost rather than their individual cost. At SO, it will be likely that travelers can unilaterally change routes to reduce their individual costs, meaning that the pattern will be difficult to sustain without some external control mechanism (Fernandez and Friesz 1983; Sheffi 1985).

The difference between UO-S and SO is clear when one considers the type of information required for travelers to achieve each pattern. UO-S postulates that travelers consider the average cost on routes: travelers choose the route between an O-D pair that has the minimum average cost for their user class (13-1), (13-2). In order to obtain the SO pattern, travelers only consider the *marginal costs* for routes, that is, the *added cost of their entry* into a route. The SO conditions imply that, at equilibrium, flow only occurs along routes whose marginal cost for the mode is minimum for that O-D pair (13-6), (13-7). Thus, travelers will only choose routes that minimize their impact on total travel cost.

SO-based model formulations have two valuable features. First, SO flow patterns provide a valuable benchmark for assessing the efficiency of other flow patterns (Sheffi 1985). In addition, while the SO principle has traditionally been viewed as an unrealistic ideal, the increasing popularity and sophistication of congestion pricing policies and ITS in general can make these conditions an obtainable goal for real-world settings.

7.3.1.4 Stochastic user optimal (SUO)

The *stochastic user optimal* (SUO) is a relaxation of a strict behavioral assumption implied by UO. In particular, SUO assumes cost minimization but allows cost *perceptions* to vary among travelers. The SUO principle is (Daganzo and Sheffi 1977):

> (SUO) At network equilibrium, no traveler can reduce his or her *perceived* travel costs by unilaterally changing routes.
>
> Alternatively: no traveler *believes* he or she can reduce costs by unilaterally changing routes.

The SUO principle assumes that the route travel costs include random components that reflect variations in travelers' perceptions. Randomness results from factors such as limited information, decision making inaccuracies or non-measured route attributes (Daganzo and Sheffi 1977). Although random variables, travel costs are related in a systematic and rational manner to the *actual* travel costs; specifically, the random travel costs result from an "error" distribution around the actual route cost. The error has an expected value of zero, meaning that the expected value of the random route cost is equal to the actual cost (13-10),(13-11). Thus, we expect perceived route costs overall to be accurate but allow for variations in accuracy across travelers

The SUO conditions require a dispersed allocation of the flow between an O-D pair according to the *probability* that each route is cheapest for travelers (13-8), (13-9). Different assumed probability distributions for the error component result in different analytical models for calculating the route choice probabilities (e.g., Sheffi and Powell 1981, 1982). However, at equilibrium the *actual* route costs for used routes will not be equal and minimal as in the UO case (Sheffi 1985). In general, under SUO each route between an O-D pair will have a non-zero flow level, although it may be small in some cases.

Although SUO has a realistic behavioral foundation, it is not as widely used as the UO principle in model formulations. This is due to the *route enumeration problem*. Calculating route choice probabilities generally requires specifying each possible route between an O-D pair: this set can be extremely large. SUO can be solved by identifying a subnetwork of likely routes rather than using all possible routes between an O-D pair, although this introduces some error (e.g., Damberg, Lundgren and Patriksson 1996; Dial 1971). In addition, the inherent nature of the stochastic flow pattern makes it difficult to search for an optimal solution to the model (Sheffi 1985). Finally, under highly congested conditions the SUO pattern closely resembles the UO-S pattern. As the network becomes congested, the equilibrium effects become stronger than the route dispersion effects due to the stochastic route choice component and the SUO solution begins to resemble the UO solution (see Sheffi 1985, 336-338 for a clear and intuitive demonstration). Nevertheless, recent breakthroughs in SUO techniques are making this theoretically appealing approach more viable from a practical perspective (e.g., Leurent 1995). Some of these techniques will be discussed below.

7.3.2 *Market equilibrium and the shortcomings of the four-step approach*

The "higher-level" demand patterns for TG, TD and MS are linked very tightly to the equilibrium pattern at the network-level. This results from these demands being elastic (that is, responsive) to the network flow costs. For example, the flow generated from origins can be influenced by travel costs since travelers may postpone or substitute other activities (e.g., telecommuting or teleshopping) when costs are high. Similarly, the amount of flow attracted to a destination can affect its attractiveness, i.e., greater congestion makes a destination less attractive to travelers. The amount of flow on the street network will decrease if travel costs are high and travelers switch to other modes In turn, postponing trips and switching to other destinations or modes reduces the network flow levels and therefore can lower travel costs.

At market equilibrium, the travel pattern should exhibit stability that *simultaneously* encompasses all four of the travel demand components. For example, at a UO-type market equilibrium, no traveler should be able to unilaterally change his or her trip propensity (TG), destination choice (TD), modal choice (MS) nor route choice (NA) without incurring higher costs. As noted above, since these components are tightly linked it is impossible to solve for each component in isolation without considering its effects on the other components. (Note, however, that *empirical* measurement of linkages between daily trip generation rates and other travel demands has proven to be problematical; see Southworth 1995).

The tight interconnections among the different travel demand components are clear when examining the formal conditions for market equilibrium given a UO-S network equilibrium. As with a UO-S network equilibrium, we identify the minimum travel cost between an O-D pair for each mode (13-13) and only allow positive flow levels on routes that exhibit that minimal cost (13-12). We also require the summed route flows for each user class between an O-D pair to equal the total travel demand for that O-D pair (13-14). Similarly, route costs are also required to be the sum of the costs for the arcs that comprise each route (13-15). Finally, all route flows and minimum costs must be non-negative (13-16). However, unlike the UO-S conditions, the O-D travel demands are no longer fixed and exogenous but dependent on the minimum route costs between the O-D pair. The summed route flows between an O-D for a mode now must equal an aggregate travel demand level determined by the travel costs between that pair (13-14). Since these costs in turn depend on route flows, we have a "Gordian Knot" of intertwined influences that must be met simultaneously.

The functional dependencies among the different travel demand components in the market equilibrium conditions *requires* any travel demand model to link *in a theoretically consistent manner* the different travel demands, their influences on travel costs and the influence of these costs on demands. Without these explicit linkages, the model does not meet the market equilibrium requirements and is consequently misspecified (Aashtiani and Magnanti 1981; Fernandez and Friesz 1983). The traditional, four-step approach violates these market equilibrium conditions (or, more correctly, *does not guarantee* these conditions) since it does not contain theoretically consistent links among the components nor an explicit mechanism for satisfying the equilibrium conditions simultaneously across all components. In contrast, convergence is the very essence of the equilibrium approach and is central to the solutions generated by these models (Boyce, Zhang and Lupa 1994).

Several studies have demonstrated the weakness of the four-step approach. As far back as the mid-seventies, Florian, Nguyen and Ferland (1975) determined that sequential estimation with feedback loops of TD⧖NA does not converge to a consistent solution. More recently, a study by COMSIS Corporation (COMSIS 1996) compared the four-step approach without feedback to the same approach with several different feedback mechanisms and a theoretically-consistent network-equilibrium approach, specifically, the Evans (1976) algorithm (referred to as the "method of optimal weighting" in the report). The "direct (feedback) method" did not consistently converge to an equilibrium solution. Feedback mechanisms based on the *method of successive averages* (MSA) compared favorably with Evans (1976) model with respect to convergence results, although the study recognizes that the MSA-based approach may not perform as well in large networks with high levels of congestion.

An extensive analysis by Boyce, Zhang and Lupa (1994) compared the four-step procedure, with and without feedback loops, with the Evans (1976) model. Specifically, the methods compared: i) one iteration through the TD⧖MS⧖NA with an "all-or-nothing" (AON) network assignment; ii) multiple iterations through TD⧖MS⧖NA with AON assignment; iii) multiple iterations through TD⧖MS⧖NA with AON assignment and MSA applied at each iteration; iv) multiple iterations through TD⧖MS⧖NA with UO-S assignment and MSA; and, v) the Evans (1976) algorithm. In many respects, the COMSIS (1996) report is similar to this study, although Boyce, Zhang and Lupa (1994) conclusions are more negative with respect to the four-step/feedback alternatives to the network equilibrium-based approach. The Evans (1976) algorithm was superior in reproducing known data, particularly key variables such as automobile link flows and total automobile trips, with only modest increases in computational effort compared with the four-step/feedback loop alternatives.

Boyce, Zhang and Lupa (1994) conclude their research paper with several recommendations that are relevant to the objectives of this current report. These recommendations are:

i) Progress in improving travel forecasts will not result from solving the four-step approach with feedback. Rather, progress will be achieved when professional practitioners begin to understand the requirements of the desired equilibrium solutions;

ii) Practitioners should insist that software vendors correctly implement methods for achieving equilibrium solutions;

iii) Federal agencies such as FHWA should conduct short courses to introduce practitioners to equilibrium-based approaches;

iv) University instructors and textbook authors should update their courses and instructional material to produce a new generation of professionals who understand the principles of equilibrium travel models.

This research report attempts to meet some of the Boyce, Zhang and Lupa (1994) recommendations by providing an accessible review of transportation equilibrium theory and practical models within that theory. The next section of this report reviews some practical equilibrium-based travel demand models.

8. EQUILIBRIUM TRAVEL DEMAND MODELS

8.1 Overview

This section of the report provides an overview of selected models that determine market equilibrium travel demands for a study area. The discussion classifies models according to the network equilibrium assumed. Note that the previous section only identified the *theoretical* conditions for network and market equilibrium. This section discusses *practical* models whose solutions correspond to the theoretical conditions discussed previously. Table 8-1 provides an overview of equilibrium travel demand models reviewed in this section.

	Travel demand components			
Network equilibrium class	NA	NA/MS	NA/MS/TD	NA/MS/TD/TG
UO-S	Sheffi (1985)	Evans (1976)	Florian and Nguyen (1978)	STEM (Safwat and Magnanti 1988)
UO-G	T2 (Dial 1995b)	Dafermos (1980)		Dafermos (1982)
DUO	Janson (1991a, 1991b)			
SUO	Fisk (1980)	Super- and hyper-networks (Sheffi and Daganzo 1980)		
UO-S/SUO	Trip consumer approach (Oppenheim 1995)			

Table 8-1: Overview of equilibrium travel demand models

The travel demand models discussed below generally follow the *equivalent optimization approach* first pioneered by Beckmann (Beckmann, McGuire and Winsten 1956). It is impossible to overstate the impact and importance of this initial work: Beckmann and colleagues single-handedly launched the entire field of network equilibrium-based travel demand modeling. All subsequent work in the static and dynamic equilibrium realms can trace their origins to this research.

The basic equivalent optimization strategy is to first specify a combined travel demand model, i.e., the combined NA/MS/TD/TG components. Then, an equivalent optimization problem is formed such that its solution corresponds to a market equilibrium of the combined travel demand components stated in the initial model. Typically, this problem contains a objective function to be minimized and constraints that represent flow and aggregate demand feasibility requirements. The objective function typically corresponds to some type of cost function, meaning that the resulting market equilibrium is the minimal cost travel pattern subject to the assumed equilibrium conditions.

In order for the equivalent optimization problem to correspond to a travel demand market equilibrium, we must be assured that its solution is *unique* and *equivalent* to the desired theoretical equilibrium conditions. Equivalency can be assured by comparing the "first-order" (first derivative) conditions for optima to the theoretical conditions for equilibrium. Discussion of these conditions is beyond the scope of this report. Sheffi (1985) provides a basic discussion of these conditions. Boyce (1984) and Boyce, LeBlanc and Chon (1988) review equivalency conditions with respect to particular travel demand models.

Solution uniqueness can be assured if the objective function to be minimized is *convex*. This condition can be visualized roughly by imagining a "u-shape" in two-dimensions or "bowl-shape" in three dimensions. The required shape is analogous for solution spaces in higher dimensions (where the number of dimensions is equal to the number of variables), although more difficult to visualize. For a more rigorous definition of convexity, see Varian (1992). To ensure convexity of the objective function, we must impose constraints on the travel demand components, particularly with respect to the arc flow cost functions and the demand functions. The discussion below will identify the assumptions required for each model.

In the interest of brevity and due to the pragmatic orientation of this report, this section only discusses the equivalent optimization problems. The travel demand components corresponding to these optimization problems are not discussed directly. This does not limit greatly the understanding of the models and their requirements from a practical perspective: data inputs, parameter estimation and solution procedures can still be readily identified.

8.2 Model Descriptions

8.2.1 *UO-S-based approaches*

8.2.1.1 NA (Sheffi 1985)

8.2.1.1.1 Assumptions

i) one mode (although multi-modal extensions are possible);

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) O-D flows are fixed and exogenous.

8.2.1.1.2 Model structure

The basic UO-S NA equivalent optimization problem is originally due to Beckmann, although the excellent and unfortunately out-of-print text by Sheffi (1985) provides a very clear and accessible statement. UO-S NA optimization problem has a straightforward structure, although Sheffi (1985) argues that the objective function does not have an intuitive economic or behavioral interpretation. The objective function consists of a single component: the summed cumulative costs of each arc cost function in the network given its current flow (14-1). Minimizing the sum of these flow costs across all arcs corresponds to the theoretical conditions that each traveler is on a path that is tied for the minimum cost between the O-D pair.

The decision variables in the optimization problem are the flow levels on each arc; the objective is to choose these flows so that the objective function is minimized. The arc flow levels are subject to the following constraints: i) the total flow on arc must equal to the summed flow for all paths that use that arc (14-2); ii) the flow on all routes between an O-D pair must sum to the aggregate travel demand for that pair (14-3); and, iii) all path flows must be non-negative (14-4). While the objective function is stated in terms of arc flows, the constraints are stated in terms of path flows. These are related to each other through an arc-path incidence variable (12-9).

The objective function of the UO-S NA optimization problem is convex and therefore has a single minimum point. Convexity is ensured by the assumptions of separable, non-negative and increasing arc cost functions. While the non-negativity and increasing assumptions are reasonable, the separable cost function assumption is restrictive as noted above.

8.2.1.1.3 Data requirements and parameter estimation

The data required for solving the UO-S NA are: i) a transportation network and O-D zonation system for the study area; ii) the aggregate flows between each O-D pair (this also implies that the total outflow from origins and the total inflows to destinations are known); iii) arc flow cost functions that meet the constraints specified above (e.g., (7-1), (7-2)); and, iv) the estimated parameters for the arc flow cost functions. Of these data/parameter requirements, only iv) presents major difficulties.

Estimating the VOT parameter in equation (7-1) is necessary if monetary expenses are relevant (e.g., tolls on certain links). This requires some type of experiment or survey in which travelers make choices among different combinations of travel times and monetary costs. Estimating the parameters of an arc performance function such as (7-2) can be expensive and time-consuming. Branston (1976) and Chp. 13 of Sheffi (1985) provide a discussion of measurement and estimation issues. More recently, Fisk (1991) extended a probabilistic traffic flow to provide link travel time/flow relationships appropriate for NA. The model parameters relate to the mean and variance of free flow time. This allows parameter estimation without the need for extensive flow/travel time observations as described in Branston (1976) and Sheffi (1985).

There is a strong underlying belief in the literature that the UO-S equilibrium conditions are the "natural" stable flow patterns that occur in real-world settings. Therefore, a great deal of attention has been directed towards the UO-S NA procedure and its use in broader travel demand models. However, as Fernandez and Friesz (1983) note, there have been few attempts to validate the UO-S NA procedure with empirical evidence. Florian and Nguyen (1976) provide one of the few empirical validations. Using empirical data from Winnipeg, they found generally good correspondence between the UO-S NA predicted flows and observed flows. However, they note that the procedure tended to overpredict arc and route travel times. They also comment on the sensitivity of the results to the arc performance function parameters and the network details (i.e., the level of network aggregation).

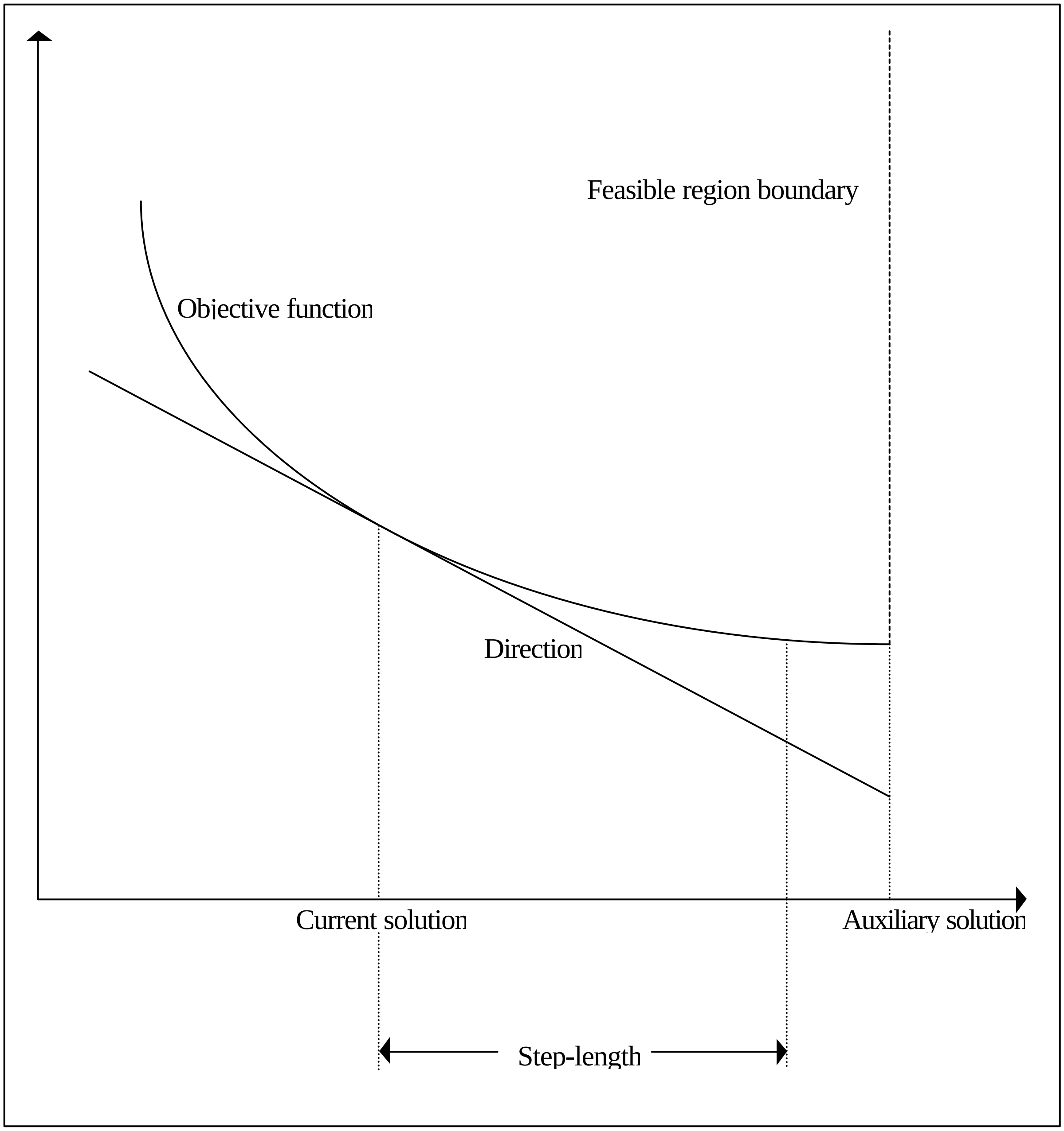
8.2.1.1.4 Solution procedure

Several heuristic solution methods have been formulated and applied to the UO-S NA problem. These include the *capacity restraint* method and the *incremental assignment* method. Capacity restraint requires a sequence of all-or-nothing assignments (i.e., all flow between an O-D pair is assigned to the shortest path between that pair) in which the previous assignment's travel costs are used for the current iteration. A problem with this approach is that the algorithm can get trapped in "cycles" where flow changes "bounce" back and forth for a subset of the network while other subsets are ignored. A "smoothing" procedure which combines the last two iteration's travel costs in a weighted average ameliorates this problem. However, neither version of the capacity restraint method converge to the desired UO-S conditions (Sheffi 1985). The incremental assignment method divides the O-D flow matrix into portions and performs an all-or-nothing assignment for each portion. After each flow portion loads on the network, the travel costs are updated and the next portion loads based on these updated costs. However, this heuristic does not converge to the desired UO-S equilibrium conditions (Sheffi 1985).

A method that generates a network flow pattern consistent with the UO-S equilibrium conditions is the *convex combinations* method (also known as the *Frank-Wolfe algorithm*). This procedure is a "feasible direction method." The algorithm starts with some feasible solution (although not an equilibrium solution, i.e., a solution that only satisfies the flow and non-negativity constraints). At each iteration, the algorithm determines the *direction* and *step-length* or move size within the solution space that best "improves" (i.e., reduces the current value) of the UO-S NA objective function. This continues until the solution can no longer be improved (Sheffi 1985). Figure 8-1 illustrates this method.

The direction-finding step of the convex combinations method involves linearizing or determining a linear approximation of the objective function at the current solution. This occurs by computing an *auxiliary solution*, i.e., an "alternative solution" based on the current solution. The two solutions form a line that determine which direction to move, i.e., which network flows and travel demands should be adjusted. However, since this line is only an approximation typically we do not want to move all the way to the auxiliary solution. Instead, we refer back to the original objective function and determine the optimal step size. This requires solving for the move-size that minimizes the objective function in that direction. Since this is a single parameter, we can easily solve for this value using a one-dimensional search algorithm. See Sheffi (1985, Chp. 4) for an excellent discussion of this strategy.

Figure 8-1: Convex combinations method (after Sheffi 1985)



The major computational effort in the convex combination method is computing the auxiliary solution. The auxiliary solution is an all-or-nothing assignment of O-D flow on the shortest path between the O-D pair based on the current flow levels (which are considered fixed during the given iteration). Therefore the algorithm must solve the set of shortest paths from each origin to all destinations during each iteration.

A problem with the convex combinations method is its slow convergence. As the algorithm nears optimum the step sizes decrease. Some improvements have been suggested for speeding-up convergence, although they do not seem to be widely implemented (Ortδzar and Willimsen 1990). Recently, Jayakrishnan *et al*. (1994) proposed a more efficient *gradient projection algorithm* for the UO-S NA problem. Its basic structure is very similar to convex combinations: the algorithm consists of a direction-finding and move-size steps. In contrast with convex combinations, the direction-finding step uses a non-linear approximation of the objective function in the neighborhood of the current solution. This speeds convergence as the algorithm approaches the optimum.

The convex combinations methods is not limited to UO-S NA: this algorithm can be applied widely to network equilibrium-based travel demand models. Adapting convex combinations to other formulations generally requires modification of the direction-finding step. These modifications will be discussed below.

8.2.1.2 NA/TD (Evans 1976)

8.2.1.2.1 Assumptions

i) one mode;
ii) separable cost functions (12-59);
iii) non-negative cost functions (12-63);
iv) increasing cost functions (12-64);
v) total outflows from origins and total inflows to destinations fixed and exogenous;
vi) TD component is a separable demand function in the form of a spatial interaction ("gravity") type function with an exponential cost function (14-5);

8.2.1.2.2 Model structure

The Evans (1976) TD/NA model extends the UO-S NA model to include a TD component. O-D flows are influenced by the minimum route cost between each pair through a spatial interaction or "gravity"-type TD component. This TD component is not explicit in the equivalent optimization problem but rather is implied by the optimality conditions for that problem. Evans (1976) main contribution was to combine foundational work by Wilson (1967, 1974) on "entropy-maximizing," doubly constrained spatial interaction models within the UO-S-based NA optimization problem developed by Beckmann, McGuire and Winsten (1956). The joint optimization problem combines in a consistent manner the flow-related costs associated with the network equilibrium, a TD based on the route costs and the TD's influence on the network flow levels.

Evans (1976) NA/TD model equilibrium requires solving a constrained minimization problem similar to the UO-S-based NA problem. The objective function (14-6) consists of two components: i) a arc-flow cost component equivalent to the NA objective function; and, ii) an entropy term that corresponds to the trip distribution model. The decision variables to be solved when minimizing this function are the flow levels on each arc and the aggregate flows between each O-D pair.

The TD term of the objective function allocates flows according to entropy-maximizing principles. In brief, this requires the flow pattern to be the most likely or highest probability pattern consistent with known aggregate information about the system (see Fotheringham and O'Kelly 1989; Webber 1977). In this case, the known information include: i) total outflows from each origin; ii) total inflows to each destination; and, iii) the minimum travel costs between each O-D pair. The flow variable values that minimize the TD component of the objective function generate the most likely TD pattern given this information.

Constraints on the Evans (1976) minimization program generally correspond to standard flow totaling and non-negativity conditions. These include: i) flows on all routes between an O-D pair must sum to the total flow between that pair (14-7); ii) flows on all routes that use an arc must sum to the total flow on that arc (14-8); iii) the flows entering each destination from all origins must sum to the known total inflows to that destination (14-9); iv) outflows from each origin to all destination must sum to the known outflows from that origin (14-10); v) all path flows and aggregate O-D flows must be non-negative (14-11), (14-12).

Evans (1976) provides a rigorous proof that the TD/NA objective function is convex and therefore has a unique minimum. From an intuitive perspective, we can note that the NA component's convexity is ensured by the same arc cost function assumptions as in the NA optimization problem (separable, non-negative and increasing). Also, the TD component is a convex function. Since the sum of two convex functions is also convex, we know the overall objective function is convex.

8.2.1.2.3 Data requirements and parameter estimation

In addition to the estimation issues discussed in conjunction with the UO-S-based NA problem, we now must estimate the parameters of the TD component's cost function. These parameters relate the effect of the minimum travel cost between an O-D pair on the amount of travel flow between that pair. Evans (1976) is silent on these estimation issues. Sheffi (1985) discusses these issues in some, although not complete, detail. To examine these estimation issues, we must turn to the literature on spatial interaction models.

Two general methods exist for estimating the parameters of a spatial interaction model such as the Evans (1976) TD component, namely, *ordinary least squares* (OLS) and *maximum likelihood* (ML) estimation. In both cases, the estimation procedure requires estimates of: i) the minimum travel costs between O-D pairs; and, ii) aggregate flows between O-D pairs. Obtaining these data items can be problematic since both are the expected outcomes of the modeling exercise itself. In the former case, the analyst may need to develop a surrogate measure for the minimum travel costs since this is determined by the NA. Possible surrogate measures include: i) assuming "free flow" (i.e., uncongested) conditions and computing the shortest path between an O-D pair and its resultant travel time; or, ii) conducting a survey and asking respondents for their travel time estimates between particular O-D pairs. Both surrogates are likely to introduce error. A third possibility, discussed by Florian and Nguyen (1978), is to perform a UO-S-based NA using the known O-D matrix to obtain reasonable estimates of the O-D minimal travel costs.

Obtaining an O-D flow matrix will require either primary (survey) data, updating existing data or through some estimation procedure. Updating existing but dated O-D matrices can occur using methods such as growth factor methods, although these methods are reliable only over short-term planning horizons (see Ortδzar and Willumsen 1990). Procedures also exist for estimating O-D matrices from link flow observations. Sheffi (1985) provides a basic albeit dated discussion of these methods; other references include Cascetta and Nguyen (1988), Fisk and Boyce (1983), Nguyen (1984), Speiss (1987) and Yang, Iida and Saski (1991, 1994). Bell (1991) discusses a statistical procedure for estimating O-D matrices from combined traffic counts/survey data.

Once the required data items are in place, the analyst must choose between OLS and ML estimation of the TD parameters. Fotheringham and O'Kelly (1989) provide an accessible discussion of both estimation procedures. Sen and Smith (1995) provide a more rigorous review of OLS and ML estimation as well as performance results of different algorithms with the approaches.

OLS is the classic and commonly-known "regression" approach to estimation. This requires transformation of the non-linear TD model to a linear form. Once the model is linearized, standard statistical packages such as SPSS™ or SAS™ can be used to estimate the parameters. While straightforward, OLS estimation of the TD parameters has some problems, including difficulty dealing with zero O-D flows and misleading goodness-of-fit statistics reported by standard statistical packages when applied to a log-transformed model. These problems do not occur with ML estimation procedures. In this case, we are trying to generate the model parameters that maximize the likelihood of reproducing the observed data from a theoretical distribution. For example, with a doubly constrained spatial interaction model we can typically assume that the interactions result from a multinomial probability distribution (Batty and Mackie 1972). Then, we estimate the model parameters by determining values that maximize the likelihood that the observed O-D flow matrix would result from this theoretical distribution. Determining these parameter values requires some type of non-linear search technique such as the Newton method (again, see Fotheringham and O'Kelly (1989) or Sen and Smith (1995) for discussions). While ML estimates are more reliable, procedures for ML estimation are not as available as OLS procedures. A spatial interaction model-specific estimation package that uses ML procedures is SIMODEL (Williams and Fotheringham 1984).

As noted above in the section on "Market equilibrium and the shortcomings of the four-step approach," the Evans (1976) model has been tested empirically. In fact, it is one of the few equilibrium-based travel demand models that has been subject to empirical validation. The studies by COMSIS (1996) and Boyce, Zhang and Lupa (1994) demonstrate the superiority of the Evans (1976) model relative to the classic four-step approach. However, more extensive empirical validation of this and other equilibrium-based travel demand models is certainly warranted.

8.2.1.2.4 Solution procedure

Evans (1976) developed a very efficient solution procedure for the TD/NA model, specifically, the *Evans partial linearization* technique. This technique is closely related to the convex combinations method; the two methods differ primarily with respect to the direction-finding step at each iteration. The direction-finding step in Evans conducts only a partial linearization of the objective function at the current solution. Evans' method is only appropriate when O-D travel demands are consistent with a doubly constrained spatial interaction model (Friesz 1985).

The direction-finding step of the Evans' partial linearization technique involves computing an auxiliary solution for the O-D flows. The algorithm computes the shortest path tree from an origin to all destinations based on the current iteration's flow costs. Based on these costs an auxiliary O-D flow matrix is computed using the partially linearized objective function. In turn, the auxiliary O-D flows are used to calculate auxiliary link flows. Then, an optimal step-size routine determines the proper adjustment of the current solution.

Evans' (1976) method is generally more efficient computationally than applying the convex combinations to the same problem. Although it still requires the major computational effort of computing shortest paths from each origin, the algorithm tends to converge faster than convex combinations since it adjusts the entire O-D matrix during each iteration. In contrast, convex combinations updates only a subset of the O-D flows during each iteration. The relative advantage of Evans over convex combinations is related to the number of positive interzonal flows in the O-D matrix. The performance of the convex combinations method may be more competitive if the O-D matrix contains some zero elements (Boyce, LeBlanc and Chon 1988).

8.2.1.3 NA/MS/TD (Florian and Nguyen 1978)

8.2.1.3.1 Assumptions

i) two modes ("automobile" and "public transit");
ii) separable cost functions for automobile mode (12-59);
iii) non-negative cost functions for automobile mode (12-63);
iv) increasing cost functions for automobile mode (12-64);
v) public transit arc costs are fixed and exogenous;
vi) TD component is a separable demand function in the form of a spatial interaction ("gravity") type function with an exponential cost function (14-13);
vii) MS component is a binomial logit model (14-14).

8.2.1.3.2 Model structure

Florian and Nguyen (1978) combine the UO-S NA model with a combined entropy-maximizing MS/TD component (see Ortδzar and Willumsen 1990). The MS/TD component combines a binomial logit model (MS) with a doubly constrained spatial interaction model (TD). The two models share the same parameter to control the cost function effect in the spatial interaction model as well as the modal split dispersion.

The objective function in the Florian and Nguyen (1978) consists of three components (14-15): i) an entropy component that determines TD and MS for the automobile mode; ii) a modified entropy component that determined TD and MS for the public transit mode; and, iii) the standard UO-S NA cost component. The modification of the public transit entropy component accounts for the fixed travel costs assumed for that mode. Components i) and ii) together comprise the combined TD/MS for both modes. The decisions variables to be determined when minimizing this objective function include: i) the aggregate travel demand for the automobile mode between each O-D pair; ii) the aggregate travel demand for the public transit mode between each O-D pair; iii) the route flows for the automobile mode; and, iv) the arc flows for the public transit mode.

Constraints on the Florian and Nguyen (1978) TD/MS/NA model comprise the standard aggregate travel demand constraints, albeit modified to account for the particulars of their "two modes with fixed costs for one mode" model. These constraints include: i) flows for both modes leaving a destination must sum to the known (exogenous) total outflow from that destination (14-16); ii) flows for both modes entering a destination must sum to the known (exogenous) total inflow to that destination (14-17); iii) route flows for the automobile mode between an O-D pair must sum to the aggregate automobile travel demand for that O-D pair (

14-18); iv) the total flow on an arc is equal to the automobile flows on routes that use that arc plus the public transit flow contribution to that arc (this latter quantity may be zero if routes are separated) (14-19), and; v) aggregate travel demands and route flows for both modes must be non-negative (14-20), (14-21).

The Florian and Nguyen (1978) model offers some practical advantages with respect to parameter estimation and computational tractability. However, these advantages require some theoretical costs, particularly with respect to assumptions regarding mode behavior and modal interactions. First, note that the model only allows two modes; this can be a drawback when analyzing travel demand in large urban areas with multiple modes. Second, note that travel costs (including travel time) for public transit are fixed, meaning that these costs are not affected by congestion. Thus, the model assumes that public transit travel times remain constant even when the network is highly congested. This is not a problem if the public transit mode is separate from the automobile network (e.g., subways) but can be a problem when public transit shares the automobile network. This problem is mitigated to some degree if the public transit schedules are accurate reflections of actual travel times, although these schedules may become less accurate when forecasting more congested conditions in the future. Also note that although public transit is not affected by congestion, public transit can affect automobile congestion. The total flow on an arc is comprised of the automobile flow plus any contribution made by public transit; this can be modified by flow equivalency factors (14-19).

8.2.1.3.3 Data requirements and parameter estimation

In addition to the flow cost function estimation issues discussed in conjunction with UO-S-based NA, the Florian and Nguyen (1978) model requires the estimation of a single parameter. This parameter controls both the cost function effect in the TD component as well as choice dispersion in the MS component. While convenient for estimation purposes, this requires a single parameter to serve a “double role” and introduces error into both components (see Ortδzar and Willumsen 1990). However, Florian and Nguyen (1978) also discuss the possibility of using two mode-specific parameters. This reduces the harsh informational requirements imposed on a single parameter to some degree, although each parameter is still required to control TD and MS effects for that mode.

Florian and Nguyen (1978) provide explicit discussion of parameter calibration for their model. They assume that an O-D flow matrix is available from survey or secondary sources; this provides the total outflows from origins and inflows to destinations. Their procedure requires first performing a UO-S NA. This assignment provides estimates of the O-D minimum travel cost values (i.e., (12-18)) for the automobile mode. A simple shortest path calculation within the public transit network provides the corresponding values for that mode. Then, using the entropy-maximizing principles developed by Wilson (1967, 1974), the parameter can be calculated by attempting to get the predicted weighted mean trip length to match the observed weighted mean trip length as closely as possible. Again, the literature on spatial interaction model calibration (e.g., Fotheringham and O’Kelly 1989) provides guidance.

8.2.1.3.4 Solution procedure

Florian and Nguyen (1978) formulate a very efficient solution procedure for their model. Similar to Evans (1976) partial linearization technique, their procedure is a modification of the convex combinations method. Again, the main modification concerns the direction-finding step.

The Florian and Nguyen (1978) reformulates the direction-finding step as a modified *Hitchcock transportation problem*, a special case linear programming (LP) problem that distributes flows between O-D pairs based on fixed arc costs. The modified Hitchcock LP determines an auxiliary solution in terms of the mode-specific O-D flows. After solving the LP, auxiliary link flows are calculated based on assigning the automobile O-D flow along the previously calculated shortest path. If the auxiliary solution has not converged with the current solution, an optimal step-size calculation occurs and the flow is adjusted for the next iteration.

The Florian and Nguyen (1978) algorithm still requires computing the shortest path trees from an origin to all destinations. This determines the shortest path cost for the automobile mode from the origin to each destination based on the current iteration's flow costs. These costs are used to initialize some of the auxiliary O-D flow variables for input into the modified Hitchcock LP. This initialization allows conversion of a two-mode version of the Hitchcock LP to an equivalent one-mode problem. This results in substantial computational savings in an already efficient LP problem. Nevertheless, the algorithm's computational effort is dominated by the shortest path calculations so the major question concerns the number of iterations required for convergence. Neither computational nor analytical results regarding this issue are provided by the authors.

8.2.1.4 TG/TD/MS/NA - STEM (Safwat and Magnanti 1988)

8.2.1.4.1 Assumptions

i) a separate subnetwork represents each transportation mode in the study area (14-22);
ii) separable cost functions (12-59);
iii) non-negative cost functions (12-63);
iv) increasing cost functions (12-64);
v) TD component is a separable demand function (12-61) in the format of a logit model whose utility function consists of the minimum travel cost between the O-D pair and a non-transportation-related destination attractiveness measure (14-23);
vi) TG is a linear function of each origin's accessibility to destinations and other, non-transportation relative "propulsiveness" factors (14-24).

8.2.1.4.2 Model structure

The simultaneous transportation equilibrium model (STEM) encompasses all four components of a travel demand analysis (Safwat 1987a; Safwat 1987b; Safwat and Magnanti 1987; Safwat and Walton 1988). The STEM objective function combines the UO-S NA component with two entropy components, specifically a TD and TG component. STEM incorporates MS by assuming that separate subnetworks represent each mode in the study area. Therefore, the UO-S paths though the overall multimodal network are simultaneous MS/NA for travelers. The disadvantage of this approach is that STEM represents modal choice as a deterministic process; this conflicts with STEM's representation of the TG and TD decisions as stochastic (see below). An advantage of this approach is it can accommodate mixed-mode trips, e.g., "park and ride" transit situations.

STEM formulates the TG and TD components through a random utility decision process at the individual traveler level. The observed utility component consists of two variables: i) the minimum average travel cost between the O-D pair (12-18); and, ii) a composite variable reflecting the non-transportation-related attractiveness of that destination (14-27). The destination attractiveness composite variable is exogenous; this can be the result of an external, separate model (e.g., a regression analysis of inflows against variables such as the amount of retail or office space). The travel cost variable has an associated negative parameter to reflect the disutility of travel. The unobserved or random utility component is assumed to have a "type I extreme value distribution," in other words, the typical error assumption used to derive a logit choice model. Some additional comments regarding this assumption are below.

The TG component generates flow from origins based on two factors: i) a composite variable that takes into account non-transportation-related factors on origin outflows (14-26); and, ii) the accessibility provided to that origin by the transportation system (14-25). Similar to the destination attractiveness composite variable, the origin composite variable is exogenous and can result from an external model (e.g., a regression model of the observed trips against residential population density in the particular origin). The second TG component measures the "accessibility" as the *expected maximum utility* of that origin. The "expected maximum utility" measures the benefit of travel from the origin assuming random utility-maximizing decisions. The accessibility variable can assume any positive or negative value; however, the STEM equivalent optimization program includes a constraint that requires this variable to assume non-negative values (14-32) since negative accessibility (and negative origin outflows) are nonsensical. The first component of the STEM equivalent optimization program's objective function (14-28) reflects the TG theoretical basis at the aggregate level.

The TD component uses the utility function to distribute flows generated from an origin among the destinations. Logit model-generated destination choice probabilities are multiplied by the number of travelers leaving each origin to estimate the flow from the origin to each destination (14-23). The second component of the STEM objective function generates entropy-maximizing O-D flow estimates consistent with the logit TD model.

The logit-based foundation of the TG and TD STEM components has both strengths and weaknesses. A strong aspect of the logit foundation is its robustness and tractability. With respect to robustness, Safwat and Magnanti (1988) demonstrate that STEM can approximate any doubly constrained spatial interaction model with fixed and known origin outflows and destination inflows. This occurs by defining the origin propulsivesness variable (14-26) and the destination attractiveness variable (14-27) as functions of the known outflows and inflows (respectively) and by restricting certain STEM parameter values (see Safwat and Magnanti 1988, Appendix B). Thus, STEM can accommodate a wide range of data for defining factors that affect TG and TD. This can allow the model to adapt to changes in available data and relevant policy variables. With respect to tractability, the logit choice model only requires very basic calculations and therefore can be applied to very large choice problems without undue computational burden.

The major weakness of STEM's logit foundation are theoretical problems related to the *Independence from Irrelevant Alternatives* (IIA) property (see Wrigley 1985). This property implies that the ratio of choice probabilities for any two alternatives should *not* depend on any other alternatives available to the decision maker. The IIA property means that the logit TD model is misspecified since it cannot account for the spatial context of the destinations. This is due to the logit assumption that choice errors among alternatives are independent (Wrigley 1985).

A simple example of the IIA property follows. Assume a decision maker is faced with two alternatives. If a third alternative is added to the choice set, then the ratio of logit-based choice probabilities between the original two alternatives will remain the same (although the *absolute* choice probabilities for both will decrease). Intuitively, this means that the third alternative will draw patronage equally from both of the original alternatives. This property is problematic when alternatives are *related*, that is, they share some attributes in common. For example, if the original two alternatives are a "central city" and a "suburban" shopping destination and a new "suburban" destination is added, we would expect the new alternative to draw proportionately more from the original "suburban" alternative than from the "central city" destination due to their shared attributes (Wrigley 1985). Fotheringham (1986) discusses a more general, spatial effect: individuals use a hierarchical information-processing heuristic that clusters proximal destinations. A logit-based TD component cannot capture these effects.

8.2.1.4.3 Data requirements and parameter estimation

In addition to the transportation network data requirements inherent in estimating UO-S equilibrium, STEM requires information on origin propulsiveness and destination attractiveness factors. As noted above, STEM is extremely flexible in this respect. The composite variables associated with the origin and destinations can be the result of independent, external models that relate factors such as (say) land use, population density, office or sales square footage to origin outflows and destination inflows. This allows STEM to be linked to land use and population forecasts to predict impacts on traffic congestion, modal split and transportation system characteristics. In addition, as noted above STEM can also accommodate known origin outflow and destination inflow information.

STEM contains two parameters that require estimation (in addition to the link performance function and value-of-time parameters associated with the transportation network). These are: i) an *accessibility parameter* that relates transportation system performance to the number of trips generated from origins; ii) a *travel cost disutility* parameter that measures the sensitivity of travel utility between O-D pairs to their travel costs. Safwat and Magnanti (1988) do not develop a statistical distribution theory for STEM that would allow efficient simultaneous estimation of both parameters. However, an estimation procedure could estimate each parameter independently. Standard procedures for estimating linear utility functions within logit models can be employed to estimate the travel cost disutility parameter (see Wrigley 1985); this requires estimates of minimum travel costs between O-D pairs as well as an observed O-D flow matrix. Estimating the accessibility parameter requires observations of origin outflows relative to minimum travel costs from that origin. Following the suggestion of Florian and Nyguen (1978), the minimum travel costs for both estimation tasks could be established by conducting a UO-S NA using the observed O-D flows.

8.2.1.4.4 Solution procedure

Safwat and Walton (1988) discuss two solution procedures for the STEM. The first procedure, the *shortest path to most needy destination* (SPND) algorithm, is an extension of the convex combinations approach. Like convex combinations, SPND determines a feasible direction at each iteration through a local linearization of the objective function. The *logit distribution of trips* (LDT) algorithm is an extension of the Evans (1976) partial linearization technqiue. Similar to the Evans (1976) algorithm, LDT uses a partial linearized objective function to update O-D flows in a more dispersed manner than the fully linearization SPND approach; consequently, its convergence is faster.

The LDT algorithm is very similar structurally to the Evans (1976) partial linearization algorithm. The algorithm first updates link costs and then calculates the shortest path tree from each origin to all destinations. Based on the shortest paths costs between O-D pairs, the algorithm allocates flows according to the logit TD function. This continues until solution convergence

Computational experience with the LDT algorithm indicates that it can be used to solve combined NA/MS/TD/TG demands on large (i.e., urban-scale) networks in a reasonable amount of time. Safwat and Walton (1988) report an application of the SPND and LDT algorithms to the Austin, Texas transportation network (3555 links, 2137 nodes, 598 traffic analysis zones and 271,625 O-D pairs). As expected, the LDT algorithm converged much quicker than SPND. Total CPU time (including generating an initial solution) was 4,734 seconds (79 minutes) on an IBM 4381 mainframe. This can be improved by a more contemporary computational platform (including current high-end desktop platforms).

8.2.2 *UO-G-based approaches*

8.2.2.1 T2-NA (Dial 1995b, 1996)

8.2.2.1.1 Assumptions

i) one mode;
ii) separable and linear cost functions (14-35) with the following components:
  a) a flow-related, deterministic component (12-40);
  b) a flow-related, stochastically-weighted component (12-41);
  c) a stochastic weight capturing varying reactions among travelers to the stochastically-weighted component (12-42). This weight has a fixed, O-D specific probability distribution (12-43).
iii) O-D flows fixed and exogenous.

8.2.2.1.2 Model structure

Many network equilibrium models include an arc generalized cost (GC) function such as (7-1) to capture the effects of travel time and out-of-pocket expense on route and mode choice. These functions equate time and money through a scalar value-of-time (VOT) parameter estimated from survey data. However, a single summary value is a poor reflection of a complex reality in which VOTs vary among individual travelers. A large amount of information is lost by using a single VOT parameter: theoretically, the VOT should be a random or *stochastic* variable to fit better variations in the population. Ben-Akiva, Bolduc and Bradley (1993) illustrate this information loss: a logit choice model's goodness-of-fit increased substantially with a stochastic VOT parameter relative to the traditional scalar parameter.

The observations regarding varying tradeoffs between travel time and out-of-pocket expense can be applied more generally. For example, travelers can have different information regarding congestion-induced delay times: travelers may be perfectly informed or may guess optimistically or pessimistically about congestion effects on routes. This involves varying tradeoffs between "known" free-flow arc times versus "unknown" delay effects. Similarly, travelers may have different risk attitudes when considering variability in travel times: this involves varying tradeoffs between average arc travel times and their variances (Dial 1995b). Both cases require a stochastic parameter in the arc GC function to adequately capture varying tradeoffs in the traveler population.

Dial (1995b, 1996) recently formulated the formal conditions and computational procedures for network equilibrium when arc GC functions have stochastic parameters. The arc GC functions contain two components, namely, a deterministic disutility (*d-disutility*) and a stochastic disutility (*s-disutility*) with a stochastic weight (*s-weight*). These components must combine in an additive manner. Given these arc cost functions, the resultant equilibrium conditions are a straightforward extension of the UO-S conditions:

> (UO-T2) Given a linear GC cost function with a stochastic weight or "s-weight," at equilibrium no travelers with his or her particular s-weight has another path with a smaller GC.

This extension mirrors Smith's (1979) extension of UO-S to the UO-G conditions that allow variety in traveler behavior while still requiring a stable, minimal cost pattern in the aggregate.

Under the UO-T2 conditions, the cost function must be minimal *for every traveler given their particular s-weight* (e.g., VOT). At first glance, this problem appears intractable. However, since the cost function (14-35) is linear only a *subset* of available paths between any O-D pair will minimize GC for *any* VOT value. This reduced set makes the UO-T2 equilibrium problem tractable. Similar to linear programming, the subset of paths that will minimize (14-35) for any s-weight are a small set of *extreme points* in cost-time space. Connecting these extreme points forms an *efficient frontier* (EF) that facilities calculation of path choice probabilities and therefore the equilibrium network flow loading.

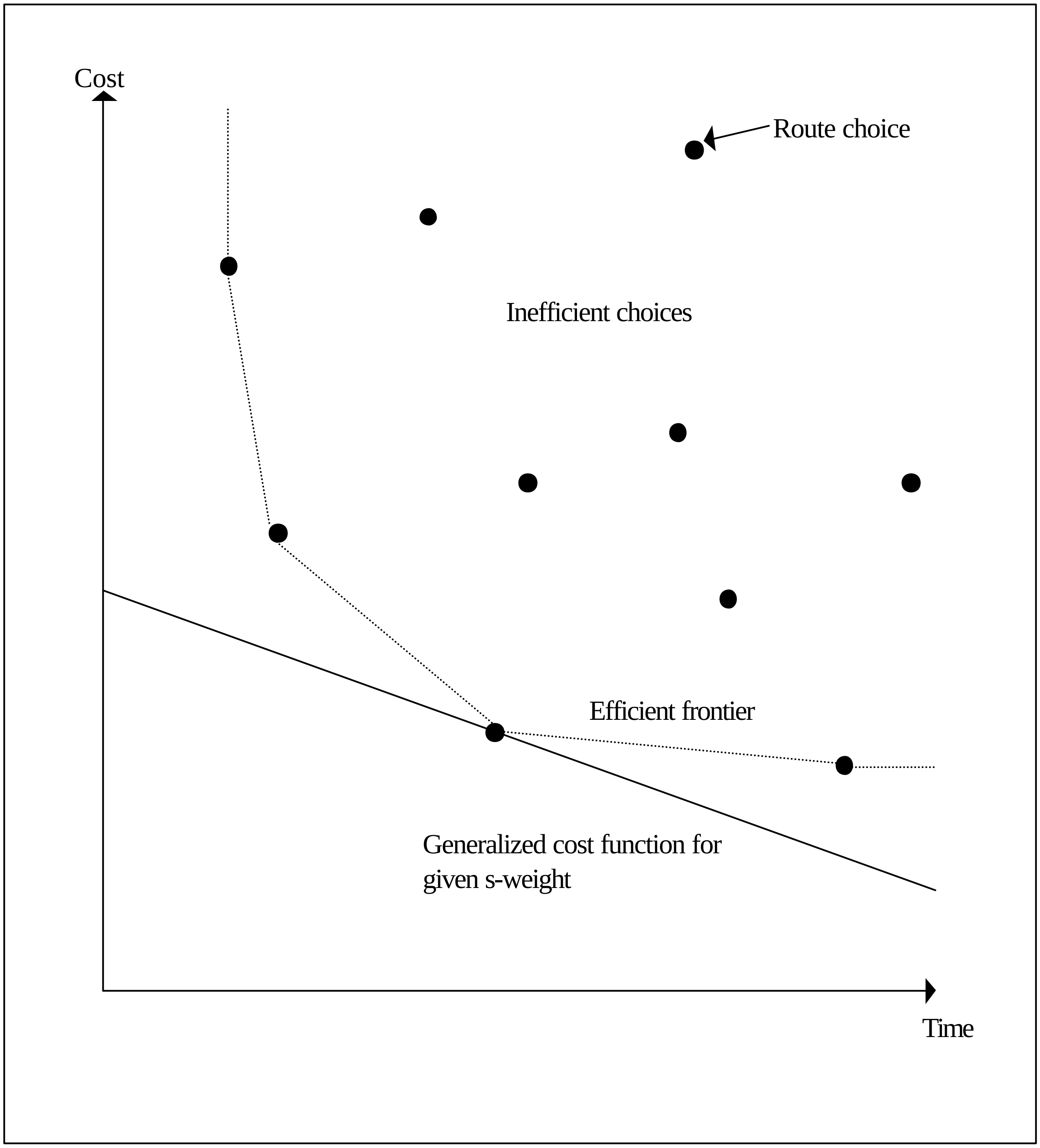
Figure 8-2 provides an example of the EF and generalized cost equation in cost-time space. A given monetary cost and travel time characterizes each path between an O-D pair and hence provides a "location" in cost-time. Only paths that comprise the lower "boundary" or EF are rational since paths above that boundary are inferior with respect to cost, time or both. The EF not only limits the number of network paths considered, but also simplifies the path choice probability calculations. Note that a particular s-weight determines the slope of (14-35) and therefore which path along the EF is optimal for that traveler (i.e., the lowest tangent between (14-35) and the EF). Therefore, the probability that a traveler will use a particular path is the probability that their VOT slope will make that path optimal. Similar to linear programming, we can calculate that probability for a given path by only considering its neighbors on the EF. We can do this in turn for each EF path to determine the proportional loading of travelers given the assumed or estimated VOT distribution.

Similar to UO-G, the formal UO-T2 conditions are also in the format of a variational inequality problem (14-36) - (14-37). This states that a flow pattern is UO-T2 if any other flow pattern would result in higher total costs, given each travelers' s-weight.

Dial (1995a) also developed a variation on T2-NA, namely, *T2-tolls*. T2-tolls defines the arc cost function (14-35) in terms of a deterministic time component and a stochastically-weighted monetary cost component. Given a set of O-D specific VOT PDFs, T2-tolls determines the toll structure for arcs that results in a SO equilibrium. As noted above in the discussion on network equilibrium theory, a SO equilibrium is ideal since it minimizes cost for the entire traveler population as a whole. However, it is difficult to achieve in practice since it requires travelers to consider their marginal impacts on congestion. The T2-tolls procedure determines a pricing system that determines, based on each traveler's VOT, the marginal social costs of congestion. When these are charged as link-based tolls, the resultant traffic flow is a

SO equilibrium. Since solving for this equilibrium toll structure is very efficient (see below), the T2-tolls procedure could be a very practical and effective congestion pricing tool.

Figure 8-2: Example T2 efficient frontier



8.2.2.1.3 Data requirements and parameter estimation

In addition to estimating the parameters of the link general cost and performance function, a major task required is estimating the s-weight probability density function (PDF). This PDF can be in any continuous, discrete or mixed format, providing great flexibility. Nevertheless, the PDF must be estimated from empirical data on traveler's route choices.

Although Dial does not develop an estimation procedure for the s-weight PDF, he provides some guidelines for this task. Dial (1996) suggests that used routes could be paired with their empirical use probability (i.e., the proportion of travelers using route x divided by the total number of sampled travelers) and fit a cumulative density function using special statistical methods (see, e.g., Silverman 1986). As noted above, Ben-Akiva, Bolduc and Bardley (1993) have recently estimated a stochastic VOT parameter in a logit mode choice model. However, their estimation procedure assumes a lognormal distribution for the VOT rather than the general distributions allowed by T2.

8.2.2.1.4 Solution procedure

A *T2-reduced simplicial decomposition* (T2-RSD) algorithm solves the variational inequality problem (14-36) - (14-37) (Dial 1995b). T2-RSD is based on the RSD algorithm originally proposed by Lawphongpanich and Hearn (1984). In turn, RSD is based on the *simplicial decomposition* (SD) procedures developed by Von Hohenbalken (1977). (Very broadly, "simplicial" is a technical term referring to constructing entities using the simplest entity ("simplex") in a given mathematical space; see Von Hohenbalken (1977) for a more technical definition.)

SD decomposes the original optimization problem into two parts: i) a *main or master problem*; and, ii) a *minor or subproblem*. The current solution of the master problem defines the minor problem objective function. In turn, the minor problem's current solution is fed to the master problem to redefine its objective function. These are solved in sequence and repeatedly until convergence. In the RSD procedure of Lawphongpanich and Hearn (1984), the master problem is the variational inequality problem of Smith (1979) while the minor problem generates minimum cost path trees (i.e., the shortest paths from each origin to all destinations) based on current flow levels. As Lawphongpanich and Hearn (1984) note, the convex combinations algorithm is a special case of the more general RSD algorithm.

T2-RSD decomposes (14-36) - (14-37) into a flow assignment subproblem and a master problem that updates the current solution by finding the optimal combination of subproblem solutions. The subproblem loads flows based on current cost levels and the s-weight intervals of the efficient frontier. Although repeatedly building shortest path trees based on different s-weights could potentially involve a high computational burden, the tree building algorithm takes advantage of the minor differences between shortest path trees based on adjacent s-weight intervals. Thus, each successive tree is a modification of the previous tree rather than a new tree built from scratch. This makes the subproblem solution very efficient: Dial (1995b) reports solving 150 minimum path trees (600,000 minimum paths) for a 4,000 node/15,000 arc network *per second*. A linear programming (LP) embedded within a linear approximation solves the master problem; this takes advantage of extremely efficient LP solution procedures. Dial (1995b) also develops a procedure for handling memory management and overflow problems in T2-RSD. The efficient solution algorithms and memory management procedures suggests that T2 NA could be a practical analytical tool for urban-scale travel demand analysis.

8.2.2.2 Combined NA/MS (Dafermos 1980)

8.2.2.3 Assumptions

i) one or more modes;
ii) non-separable cost functions (12-60);
iii) the major influence on a mode's arc flow cost is that mode's flow within that arc (14-41);

8.2.2.4 Model structure

Dafermos (1980) presents a very general NA/MS model that relaxes the restrictive non-separable arc cost function inherent in most network equilibrium-based travel demand models. Recall that these cost functions assume that a mode's link flow cost is only influenced by that mode's flow within that link. This assumption does not recognize interactions among modes within a link nor the influence of flows within other links (e.g., cross-traffic at intersection or two-way traffic on the same street). This unrealistic assumption is necessary for model tractability within the UO-S framework.

Working within the UO-G framework, Dafermos (1980) relaxes this assumption, albeit not completely. Note that assumption iii) above requires a mode's flow within a link to be the dominant component that determines that mode's arc cost. This is reasonable from the perspective of cross-link influences; i.e., we would expect cross-traffic at intersections and traffic in the opposite direction to have less influence on congestion within a link than the traffic in that link. However, this assumption is less tenable with respect to inter-modal interactions within the same link. Nevertheless, even though inter-modal interactions within a link must be subdued this is an improvement over not capturing these interactions at all.

Dafermos (1980) states the NA/MS model as a variational inequality (VI) problem (14-42). This parallels the theoretical development by Smith (1979), although Dafermos (1980) states and analyzes the model at the more convenient link flow form instead of the equivalent path flow form. The VI requires the UO-G equilibrium to be the lowest overall arc flow cost of any other feasible arc flow pattern.

Assumptions regarding inter-link and inter-modal interactions are captured through a linear link flow cost function (14-38). This function consists of two components. The first component is the set (matrix) of all arc flows in the network pre-multiplied by a matrix reflecting the inter-modal and inter-link interactions. The second component consists of static or "base" link costs. As noted above, the flow interaction matrix must be structured so that modal flow on an arc dominates that mode's cost for that link. This ensures that the cost functions behave correctly and a unique solution exists (14-41).

8.2.2.5 Data requirements and parameter estimation

The major additional data requirement and estimation task in the Dafermos (1980) MS/NA model is estimating the inter-link and inter-modal interaction matrix in the link flow cost function (14-38). Primary data required to estimate inter-link interactions are detailed, time-stamped flow and travel time observations across the network (or a sampling of key links). Estimating interactions among different modes is more difficult; this requires detailed observations of modal flow levels and link travel times. Given a lack of primary data, analysts can make assumptions regarding these interactions; this approach is often used to derive "flow equivalency factors" to capture modal interactions.

8.2.2.5.1 Solution strategy

Dafermos (1980) presents an algorithm to solve the VI problem (14-42) given the special linear arc flow cost functions (14-38). The algorithm as presented is for a single mode problem with link interactions; extending the algorithm for multiple modes is a straightforward transformation using earlier work on multiclass transportation networks by Dafermos (1972).

The algorithm requires repeated solution of a UO-S NA problem given a special transformation of the link flow cost function. The link flow cost function includes a parameter that strongly influences the convergence speed of the algorithm. Determining the proper value of this parameter requires computing the *eigenvalues* (characteristic roots) of two large matrices (specifically, (14-41) and a function of (14-41) and the mode/link interaction matrix). This can be very complex, particularly for the large matrixes implied by a large-scale application (see Press *et al.* 1992, 456-463 for a discussion). However, Dafermos (1980) notes that the parameter can be selected using a trial and error strategy and makes some suggestions about interval bounds for the parameter.

The applicability of the Dafermos (1980) to large-scale networks is unclear. Both the complexity of certain operations (particularly, calculating eigenvalues) and the large number of iterations potentially required for convergence imply that the Dafermos (1980) model may be limited to sketch planning networks. However, this report reviewed this model for completeness as well as the possibility that a more efficient algorithm could be developed using VI tools. Continued research is required.

8.2.2.6 Combined NA/MS/TD/TG (Dafermos 1982)

8.2.2.6.1 Assumptions

i) one or more modes;

ii) non-separable cost functions (12-60);

iii) the major influence on a mode's arc flow cost is that mode's flow within that arc (14-41);

iv) non-separable demand functions (12-62);

v) the major influence on the modal flow between an O-D pair is that mode's travel costs for that O-D pair (14-46).

8.2.2.7 Model structure

The Dafermos (1982) NA/MS/TD/TG model is a direct extension of the Dafermos (1980) UO-G-based NA/MS model. In this case, the model treats the travel demand between an O-D pair as elastic instead of fixed as in Dafermos (1980). This accounts for the higher-level TD/TG demands.

The model is stated in the form of a VI (14-47). The VI consists of two components: i) the link flow cost functions; and, ii) a function that measures the travel *disutility*. Formally, travel disutility is the inverse function of the travel demand function (12-24). This measures the generalized cost (disutility) associated with each travel demand level. The VI objective function requires the combined link flow costs and travel disutilities to be the aggregate minimal cost pattern among all feasible patterns.

The VI objective function (14-47) corresponds to a generalization of UO-S market equilibrium conditions at the individual level. These conditions dictate that any route between an O-D pair exhibiting positive flow has a travel disutility equal to the route's flow cost. If a route's flow cost is greater than the travel disutility than the flow on that route must be equal to zero (14-48). Any travel demand pattern that satisfies this condition can be stated in a form similar to the VI objective function (14-49). Aggregating this statement allows the relaxed behavioral conditions pioneered by Smith (1979): only the aggregate pattern is required to be minimal rather than each individual trip.

The combined NA/MS/TD/TG model contains the linear link flow cost function from the Dafermos (1980) NA/MS model. The travel disutility functional form is analogous: it contains a matrix capturing the travel disutility interactions among O-D pairs. This matrix must be structured so that the influence of a mode's flow between an O-D pair dominates that mode's travel disutility for that pair. This ensures that the travel disutility functions behave correctly and a unique solution exists (14-46).

8.2.2.7.1 Data requirements and parameter estimation

As with the UO-G-based NA/MS model, the combined NA/MS/TD/TG model requires estimation of the mode/link interaction matrix. In addition, the travel disutility matrix must be estimated. This requires estimation of non-separable demand functions, i.e., demand functions in which the flow between an O-D pair depends on the set of minimum costs across all O-D pairs (although, as noted above, the flow-related cost for the given O-D pair is the assumed to be dominant). The spatial interaction literature does not provide guidance for estimating non-separable demand functions. However, one could access spatial interaction estimation techniques by structuring the travel disutility matrix so that off-diagonal elements are zero, i.e., consistent with separable demand functions.

8.2.2.7.2 Solution strategy

The combined NA/MS/TD/TG model's solution algorithm is a direct extension of the solution strategy for the Dafermos (1980) NA/MS model. The algorithm involves repeated solution of a UO-S NA assignment with elastic demand. The algorithm requires transformations of the link flow cost and travel disutility functions. These transformations include parameters that influence strongly the convergence speed of the algorithm. These parameters must be estimated from the eigenvalues of the mode/link interaction and the travel disutility interaction matrices. This can be difficult for large-scale travel demand analysis. A trial and error search strategy to find these parameters is also possible, and Dafermos (1982) provides guidelines on the intervals for these parameters.

Similar to the UO-G-based NA/MS model, the efficiency of this algorithm for large-scale travel demand analysis is unclear. However, it is worth discussing this model due to its generality. Continued research is required to test the algorithm for a large-scale application and perhaps improve the solution algorithm speed using related solution techniques for VI problems.

8.2.3 *DUO-based approaches*

8.2.3.1 DUO NA (Janson 1991b)

8.2.3.1.1 Assumptions

i) one mode;

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) O-D flows fixed and exogenous;

vi) the study time period divided into discrete time intervals of equivalent duration (12-31), (12-32).

8.2.3.1.2 Model structure

Janson's (1991a, 1991b) DUO NA model is a direct extension of the UO-S NA model (Sheffi 1985). The objective function extends the minimization of the cumulative arc cost function across all network arcs to minimization of these costs across all discrete time periods (14-50). Thus, we essentially extend the UO-S equivalent optimization problem across multiple, discrete time periods and require minimization of the arc cost function across these time intervals.

The DUO NA (Janson 1991b) contains both static and dynamic constraints. The static constraints are equivalent to the UO-S constraints with the added dimension of the discrete time intervals. Specifically, we require: i) flows on an arc during a given time period to be equal to the summed flows that departed during any time period on any path that uses that arc during the given time period (14-51); ii) the summed flow that departs during a given time period must sum to the known flow departure total for that time period (14-52); and, iii) route flows during any time period must be non-negative (14-53). The objective function plus these constraints is exactly equivalent to the UO-S NA equivalent optimization problem when there is only one time period.

A temporal path-arc incidence variable maintains correspondence between arcs and paths during each time interval for flows that depart during the same time interval (12-37). Note that this is a temporal extension of the static arc-path incidence variable (12-9). A key difference is that the static incidence variable is an exogenous constant while temporal arc-path incidence is a decision variable solved within the problem. In DUO, the arc composition of paths for flows that departed during a given time period cannot be predetermined since the time interval of arc use is affected by travel costs which in turn are affected by flow loadings (Janson 1991b).

The endogenous nature of arc-path incidence in DUO requires the problem to have non-linear dynamic flow constraints to ensure flow continuity. First, we require flows to only use each arc on a given path only once during each time interval (14-54). Second, we require each path to use its arcs in a temporally continuous and logical manner relative to the travel times to each arc's nodes. This is accomplished by first measuring the total travel time on a path from the origin to a given node for trips departing in a given time interval (14-55). Then, we force flow to use the arcs in a path in a temporally consistent manner. Flow can only use an arc during the interval that it reaches the from-node of that arc according to the cumulative travel time to that from-node. If the cumulative travel time to the from-node is greater than or less than the cumulative "clock time" (measured by the number of elapsed intervals times the duration of each interval), then the temporal arc-path incidence variable is forced to zero and the path cannot use that arc ((14-56) and (14-57), respectively).

Janson and colleagues have developed several extensions of the basic DUO NA model. Janson (1995) formulates a DUO NA model with known (fixed) arrival times in contrast to the known departure times assumed in the original formulation. Janson and Robles (1993) developed a DUO NA model that includes departure or arrival time choices; thus, travelers choose a departure or arrival time simultaneously with their route choices. This can allow modeling of travel timing decisions. Janson and Robles (1995) develop a "quasi-continuous" version of the model by allowing fractional (as opposed to integer) flows. This allows better representation of dynamic congestion effects such as spillback queuing effects downstream from incidents such as accidents.

8.2.3.1.3 Data requirements and parameter estimation

The major data requirement for the Janson's (1991b) DUO model is a time-specific O-D flow matrix. Ideally, this requires O-D flow data "tagged" with the time of day when each trip occurred. These data can be aggregated to the discrete time intervals of the DUO model. Since the DUO model specifies a dynamic equilibrium for flows departing within the same time interval, the critical "time stamp" is the departure time of each trip although the arrival times can also be used for model validation. In contrast, the alternative formulation in Janson (1991a) requires "time stamps" corresponding to arrival times.

If a temporal O-D matrix cannot be obtained directly from primary data it must be estimated. Janson and Southworth (1992) discuss a method that uses the dynamic traffic assignment procedure to estimate departure times from observed link traffic counts; these data are often readily available. Another, less sophisticated, option is to temporally disaggregate a daily O-D flow matrix. The simplest method is to divide the O-D matrix equally into the n daily intervals implied by the specified time duration. However, since O-D flows typically exhibit morning and daily peaks rather than an even daily distribution this approach is crude. Daily O-D flows could be distributed over the time period of interest by using daily peak profile curves; this would provide a more realistic estimate of the time-dependent O-D flows

An issue that must be addressed when implementing the DUO model are the appropriate time interval duration. Janson (1991b) suggests choosing an interval that is approximately four to five times the mean link time impedance in the study area. This minimizes flow estimate variation between intervals.

8.2.3.1.4 Solution strategy

Two solution stratgies are available for the DUO NA. Janson (1991b) formulates a heuristic strategy, the *dynamic traffic assignment* (DTA) procedure, that generates "good" (near-optimal) solutions with reasonable computational times. Janson (1991a) develops an exact (optimal) algorithm, the convergent dynamic algorithm (CDA).

DTA incrementally assigns the known flows departing during each interval to shortest paths while anticipating future link volumes. Note that in a static NA problem all flow assignment occurs at the same "time." Therefore, the procedure can simply compute the shortest paths from an origin to each destination based on the current levels of arc flows and costs (although these flows can be re-adjusted until convergence to equilibrium). In the dynamic realm, it is unknown how current and future arc volumes will be affected by assignments from other origins. Thus, after each assignment the DTA procedure must project current arc flow assignments into future time intervals. DTA projects future arc flows based on current arc flow levels, ratios of future (not yet assigned) travel demands and flows assigned in previous intervals. This assumes that reasonable estimates of future arc flows can be made by multiplying current arc flows by factors that account for travel demand levels in future time intervals. DTA uses these projections only to calculate the shortest paths; these flows are only assigned during their appropriate time intervals.

Projecting the current arc flows into future time intervals occurs through a weighted combination of arc flows assigned thus far during the currently projected interval and the final arc flows from the previously projected interval. The weight given to the current interval's flow during an origin's flow assignment is equal to the percent of total trip flows assigned to that point: this ranges from 0% during the first origin's flow assignment to 100% during the last origin's flow assignment. Since these weights depend on the order in which an origin is considered, origins are randomly selected in order to randomize the arc flow loadings.

The CDA strategy combines the convex combinations method with a linear programming approach. Convex combinations solves for a UO equilibrium with fixed node time intervals. This solution is then passed to a linear program to update node time intervals. These updated node time intervals are then passed back to the convex combinations routine for flow updating. This continues until convergence; this is measured based on the number of node time intervals changed since the last iteration.

8.2.4 *SUO-based Approaches*

8.2.4.1 SUO NA (Fisk 1980)

8.2.4.1.1 Assumptions

i) one mode;
ii) separable cost functions (12-59);
iii) non-negative cost functions (12-63);
iv) increasing cost functions (12-64);
v) O-D flows fixed and exogenous;
vi) route costs are random variables consisting of an observable component and an unobservable or random component whose expected value is zero (13-10), (13-11).

8.2.4.1.2 Model structure

The SUO NA problem loads fixed O-D flows onto a network in a manner consistent with the SUO equilibrium conditions (13-8) - (13-11). As discussed above, the stochastic component attempts to reflect limited information and subjectiveness. From a theoretical perspective, this is an improvement from the UO conditions that assume perfect information and rationality on behalf of travelers.

There are several formulations of the SUO NA. These formulations differ mainly with respect to the route choice probability calculations. Generally, there are two major route probability mechanisms. Both fall within the realm of random utility theory, meaning that travelers' utility functions for route choice are assumed to have a measured and random component. A *logit*-based network loading routine is very tractable, but has some theoretical problems. First, the logit-based loading is insensitive to network topology; this results in too much flow being allocated to overlapping routes. This is due to the logit's model IIA property: the model assumes that choices are independent and do not share attributes. A second problem is the logit reliance on travel cost differences only. This implies that the *magnitude* of the route length is ignored, e.g., a five minute travel time difference has the same effect whether the route lengths are ten versus fifteen minutes or 120 versus 125 minutes in length (see Sheffi (1985), pp. 302-305 for a clear illustration). These properties weaken the behavioral foundation of the logit-based flow pattern. Despite these behavioral weaknesses, logit-based network loading is popular due to the logit model's tractability.

Probit-based network loading assumes a very general error structure, meaning that route choice utilities can be correlated. Probit-based network loading takes into account network topology and route length magnitudes. However, behavioral realism is gained at the expense of more difficult model calculations. A closed-form (analytical) solution does not exist for the general probit model, meaning that calculations must often be obtained through Monte Carlo simulation or other, computationally-intensive methods (Sheffi 1985).

Fisk's (1980) SUO NA model uses a logit-based route choice mechanism. SUO NA adds an entropy-based route flow component to the UO NA objective function (14-58). As noted above, an aggregate-level entropy component is consistent with a random utility/spatial interaction choice mechanism at the individual-traveler level (Fotheringham and O'Kelly 1989; Oppenheim 1995). The integration of the UO-S NA arc cost component and the route choice entropy component in the Fisk (1980) objective function has some interesting properties. A parameter associated with the route choice entropy component measures travelers' sensitivity to route costs. When this parameter tends to infinity, the route choice entropy component tends to zero and a UO-S NA is obtained. Thus, the UO-S NA model is a special case of the more general SUO NA model. Conversely, when the parameter tends to zero the entropy component becomes dominant. In this case, flows are evenly dispersed among routes, i.e., travelers do not consider costs when making route choices (Damberg, Lundgren and Patrikson 1996; Fisk 1980). Fisk (1980) notes that the full IIA route choice properties only occur when the parameter value is zero; the IIA properties weaken as the parameter becomes more positive.

For all positive values of the traveler cost sensitivity parameter, the SUO NA model will generate a positive flow level for each network route regardless of its travel cost, although many of these flow levels can be quite small. Since the number of routes can be quite large, solution algorithms must either define a set of plausible or *efficient* routes (see, e.g., Dial 1971) or must work directly with arc flow levels rather than route flow levels (Damberg, Lundgren and Patrikson 1996).

Constraints on the SUO NA problem correspond to standard flow consistency and non-negativity conditions. These include: i) the summed flow on routes between an O-D pair must sum to the aggregate flow between that O-D pair (14-59); ii) the summed flows on all routes that use a particular arc must sum to the total flow on that arc (14-60); and, iii) route flows must be non-negative (14-61).

8.2.4.1.3 Data requirements and parameter estimation

In addition to calibrating link generalized cost and link performance functions, Fisk's (1980) model requires estimation of the traveler cost sensitivity parameter. Since this parameter's value is uniquely determined by the optimal network flows, it can be calibrated from observed flow levels in the network (Damberg, Lundgren and Patriksson 1996; Fisk 1980). However, this causes some difficulties since these flows are endogenous to the model (Anas 1988).

Huang (1995) developed a combined algorithm for solving and calibrating Fisk's (1980) model. The combined algorithm starts with an arbitrary parameter value and solves for the SUO flows based on that value. The algorithm then compares the predicted total network flow cost to an exogenously-determined observed value. If the predicted and observed values do not match, the algorithm increments the parameter value upward or downward (depending on the comparison) and resolves for the new SUO flows. The combined algorithm is computationally-intense since it requires repeated solution for the SUO equilibrium flows. This can be mitigated to some degree by a good initial guess for the parameter value. Also, since lower values of the parameter require the enumeration of larger number of paths, it is more efficient to start with a larger value for the parameter and allow the algorithm to work "downward" to the correct value.

8.2.4.1.4 Solution procedure

Several solution strategies have been proposed for the SUO NA problem. As a very general solution strategy, the method of successive averages (MSA) can be used with any stochastic network loading routine, i.e., logit or probit (Sheffi 1985; Sheffi and Powell 1982). This method is discussed in more detail below in the section on super and hypernetworks. Discussion in this section is limited to solution algorithms specific to Fisk's (1980) model.

Chen and Alfa (1991) develop two algorithms based on the convex combination methods and Dial's STOCH network flow loading algorithm (Dial 1971). Since the algorithm optimizes the step length during each iteration, the algorithms converge much quicker than

MSA. However, the Chen and Alfa (1991) algorithms may result in inconsistent flows and can require balancing procedures to enforce consistency (Bell *et al.*1993). Damberg, Lundgren and Patriksson (1996) developed a heuristic solution strategy that provides solutions directly in terms of path (as opposed to arc) flows.

An important task in operationalizing Fisk's (1980) model is determining the subset of paths between each O-D pair that should be considered when loading flows. Since logit-based network loading theoretically loads positive flow levels on every route between an O-D pair, we must be careful in how we restrict the extremely large number of O-D routes to a more manageable subset. Fisk (1980) discusses two methods: i) shortest path assignment; and, ii) Dial's (1971) STOCH algorithm. The former method loads flow onto the shortest path identified during each iteration. The latter method uses a logit function directly to load flows onto the set of efficient paths during each iteration. The "efficient paths" are those that only include links that bring the traveler closer to the destination and farther from the origin (i.e., a path is not efficient if it brings the traveler closer to origin during any step). The path subset selection definition can vary among SUO solution algorithms; these specifications trade-off computational efficiency versus error introduced by not considering some paths between an O-D pair.

Leurent (1995) recently refined both the Fisk (1980) minimization program as well as Dial's (1971) STOCH algorithm. The improvement to the Fisk minimization program involves a more sensitive convergence test. The improvement to STOCH provides a more stable definition of "efficient paths," allowing path calculations to occur only once rather than during each iteration. Both refinements make the Fisk SUO model more competitive to deterministic approaches with respect to computational effort.

8.2.4.2 NA/MS/TD/TG - Super- and hyper-networks (Sheffi and Daganzo 1980)

8.2.4.2.1 Assumptions

i) an expanded network represents the transportation system (14-62);
ii) the expanded nodes consist of *basic* nodes (12-48) representing the transportation network (e.g., street intersections and public transit stops) and a set of virtual or *non-basic* (12-49) nodes that represent TG and MS decisions (14-63);
iii) the expanded network arcs consist of *basic* arcs (12-50) representing the transportation network (e.g., street segments and public transit route segments) and a set of *non-basic* or *entrance/egress* arcs (12-51) that represent TG and MS decisions (14-64);
iv) flow costs are fixed for arcs in the entrance/egress network;
v) separable cost functions (12-59)for arcs in the basic network;
vi) non-negative cost functions (12-63) for arcs in the basic network;
vii) increasing cost functions(12-64) for arcs in the basic network.

8.2.4.2.2 Model structure

The supernetwork approach (Sheffi 1985; Sheffi and Daganzo 1978, 1980) extends the classic network equilibrium problem to encompass other travel decisions such as TG, TD and MS. The fundamental idea is elegant: augment the "standard" transportation network with abstract links and synthetic cost functions that represent other travel decisions. Solving for an equilibrium flow pattern for the extended network provides consistent demands across the travel components represented. Sheffi (1985) provides a rigorous treatment of this method, while Slavin (1995) provides a more accessible review.

A supernetwork consists of a *basic network* and an extended or *non-basic network* (see Sheffi 1985; Sheffi and Dagnazo 1978, 1980). Figure 8-3 provides an example for combined NA/MS. In general, the basic network corresponds to a detailed, "physical" network with deterministic flow-dependent cost functions, e.g., an urban street network. The flow pattern in this network represents a solution to the NA travel demands. Conversely, the non-basic network consists of "abstract" links with stochastic flow-independent cost functions that reflect mode, destination or trip generation choices. Consequently, the non-basic flow pattern provides the TG, TD and/or MS travel demands.

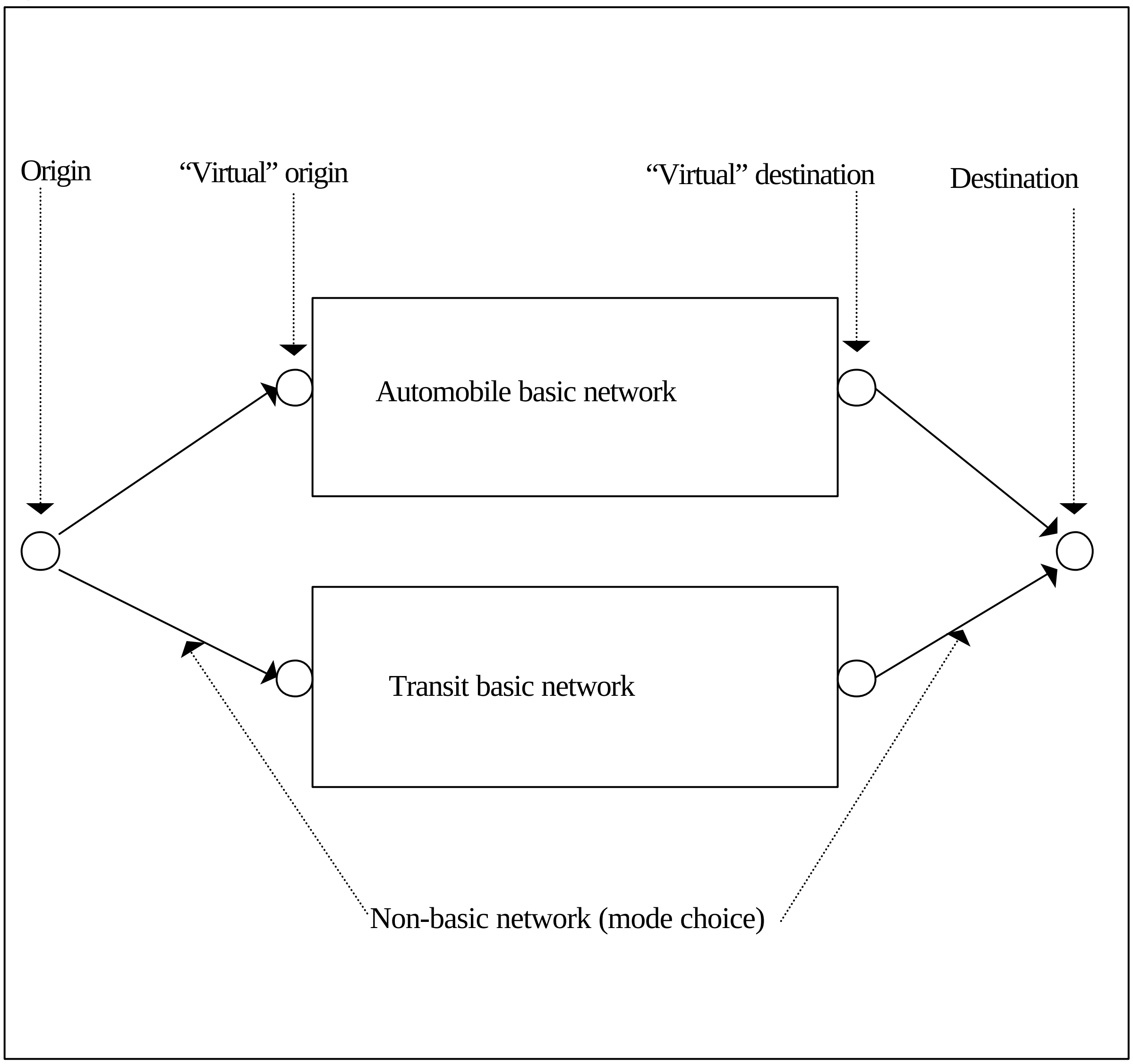
The supernetwork equilibrium conditions correspond to a SUO equilibrium across the augmented network. Specifically, these conditions require: i) the total demand between an "entrance" basic node (i.e., a basic node directly connected to one or more origins) and an "egress" basic node (i.e., a basic node directly connected to one or more destinations) is equal to the total aggregate demands between all O-D pairs connected to that basic node pair times the probability of that node pair being used (14-65); ii) the flow on all paths between a basic network entrance/egress pair must sum to the total demand between that pair (14-66); iii) positive flow levels only occur on routes that are tied for the minimum cost level between any basic network entrance/egress node pair (14-67), (14-68); and, iv) all route flow levels must be non-negative (14-69). Note that *any* stochastic loading routine can be used to calculate the usage probability for basic network entrance/egress node pairs. "*Super*network" refers to this general case while "*hyper*network" refers to the particular case of a probit choice mechanism.

The distinction between deterministic, flow-dependent costs and stochastic, flow-independent costs implies that, at equilibrium, the UO-S conditions hold for the basic network while the SUO conditions hold for the expanded network. Thus, the user equilibrium dimension derives from the basic network while the stochastic dimension results from the non-basic network. The stochastic costs on the non-basic network links relates to random utility theory and therefore is appropriate for modeling TG, TD and MS. However, the deterministic restriction on the basic network link costs is for computational tractability. In particular, this assumption allows changes in basic network flows to affect only the *average* cost between a non-basic O-D pair. This allows the choice probability for a non-basic link to depend only on the minimum impedance rather than a complete enumeration of all basic network paths between a non-basic node pair (Sheffi and Daganzo 1980).

8.2.4.2.3 Data requirements and parameter estimation

Basic data requirements and parameter estimation tasks associated with the supernetwork approach are estimating the parameters of the random utility-based demand functions for each aggregate travel demand encompassed (i.e., MS, TD and/or TG). Since the NA component is deterministic, these estimation tasks are straightforward and do not have the difficulties associated with the parameter estimation for SUO NA. The basic task is to estimate the parameters of the random utility functions. This requires observations of individual-level travel decisions and the hypothesized characteristics that influence these decisions. Since the supernetwork model is aggregate, these characteristics should be decision-specific (e.g., accessibility, level-of-service) rather than individual characteristics (e.g., household income). Sheffi (1985) provides an accessible albeit brief review of estimation procedures; a more detailed and technical discussion can be found in Ben-Akiva and Lerman (1985).

Figure 8-3: Example supernetwork for NA/MS



8.2.4.2.4 Solution procedure

Any SUO solution method can solve for this equilibrium over a supernetwork. Sheffi and Daganzo (1980) provide a method that is very similar to the UO-S convex combinations algorithm. Sheffi (1985) and Slavin (1995) discuss MSA as a good, general purpose solution algorithm that can be applied with any stochastic network loading routine.

MSA is similar to the convex combinations method. In this case, the step-size along the feasible direction is not determined during each iteration. Instead, a sequence of step-sizes are predetermined before algorithm execution. Many types of step-size rules are feasible; the only requirements are that an infinite sum of the step-sizes and the square of the step-sizes is infinity and less than infinity, respectively (i.e., the step-sizes should generally be positive but less than one). A simple step-size rule that meets this requirement is $1/n$, where n is the iteration number. The move direction is determined through a stochastic network loading model. The requirements on the move direction are also quite general: the move direction must be a descent direction only on average (Sheffi 1985). This provides a great deal of flexibility for the network loading routine (e.g., logit, probit).

MSA's fixed step length means that the algorithm requires a large number of iterations to converge (Chen and Alfa 1991; Huang 1995). Another MSA weaknesses is that its convergence is not monotonic, i.e., the flow change does not necessarily become increasingly smaller. This relates to the randomness of the move direction and the fixed move size. The convergence criteria should therefore be based on flow characteristics over several previous iterations rather than comparing the current flow with just the last iteration (Sheffi 1985).

8.2.5 *Combined UO-S/SUO Approaches*

8.2.5.1 TG/TD/MS/NA - Trip consumer approach (Oppenheim 1995)

8.2.5.1.1 Assumptions

i) one or more modes;

ii) non-negative cost functions (12-63);

iii) strictly increasing cost functions (12-64);

iv) separable cost functions (12-59) (although two-mode non-separable functions are possible);

8.2.5.1.2 Model Structure

The *trip consumer* (TC) approach (Oppenheim 1995) formulates travel demands within classic microeconomic consumer demand theory. Oppenheim (1995) obtains consistency among individual decisions and aggregate equilibrium conditions by linking the utility structures for individual travelers to corresponding aggregate-level optimization problems. Solving the optimization problem generates the equilibrium, aggregate-level travel demands corresponding to the postulated individual-level utilities. In addition, several well-known and efficient solution algorithms are available for solving the optimization problems. The TC approach is extremely broad and flexible; several established travel demand models can be derived as a special case of this general (Sheffi 1985; Evans 1976; Safwat and Magnanti 1988).

The TC approach solves the travel demand problem by restating it as an aggregate-level version of the classic *consumer utility maximization* problem in microeconomic theory (14-75) - (14-77). In this problem the consumer attempts to choose a "bundle" of goods that maximizes his or her benefit subject to a maximum expenditure limit ("budget"). At a basic level, we can view transportation services within this framework. Similar to more traditional "goods," transportation services offer benefits (i.e., accessibility to destinations) but incur costs (time, money) that travelers have varying willingness or ability to pay. A traveler chooses the type and levels of transportation services that maximizes his or her benefit subject their ability or willingness to pay. Stating the travel demand problem in this format requires: i) formulating an individual-level utility structure that encompasses the relevant travel demand components; ii) restating the individual-level choice utilities as aggregate-level utilities; and, ii) transforming the choice-specific or *indirect* utilities to *direct* utilities that, when maximized, generates the equilibrium travel demands.

In the general case, the TC approach specifies individual, choice-specific utilities within the random utility framework. As discussed previously in this report, each choice utility consists of a measured and a unmeasured component. The measured utility component can have arbitrary length and form, providing a high degree of flexibility for incorporating relevant, policy-related and behavioral factors. In contrast with the standard logit error assumption, the TC approach assumes that the that the stochastic components are independently and identically Gumbel distributed. This allows a tractable *nested logit* structure to represent interrelationships among the four travel demand utilities.

The nested logit (NL) approach is a method for representing interrelationships among choices in a random utility framework (see Wrigley 1985). The NL model assembles individual choice utilities into a composite utility structure by "nesting" expected utilities of related choices within a choice's utility function. This nesting structure often represents a sequential decision process, e.g., a NL model would reflect the temporal sequence of "Choice A then Choice B" by nesting Choice B's expected utility within the utility function of Choice A. In this case, Choice B's utility is an *expected* utility since its benefit depends on its choice probability. Another, equally valid, interpretation of the nesting structure are hypothesized *interrelationships* among the travel demand components and the relative effects of the unobserved utilities (see below). This interpretation does not assume a temporal choice sequence, although the mechanics are identical.

Appendix 14.5.1 provides the indirect and expected utilities corresponding to each travel demand component. The expected utilities provide the basis for utility nesting. For example, when modeling MS and deterministic NA, the analyst could hypothesize a nesting structure of MS/UO-S-NA. This would reflect the hypotheses that: i) route choices within a particular mode are more similar than route choices between modes; and, ii) NA has a smaller stochastic component than MS (i.e., the analyst is more certain about the NA utility function specification than the MS specification). In terms of the utility functions in Appendix 14.5.1, this would require nesting (14-87) within (14-82). Similarly, hypothesizing a TG/TD/MS/UO-S-NA nested structure would require nesting (14-85) in (14-83), (14-83) in (14-81), and (14-81) in (14-78).

The nested utility structures are used in the parameter estimation phase of the TC approach. The results of the parameter estimation phase may result in a modification of the nesting structure. After parameter estimation, a particular model within the TC framework can be solved using the direct, aggregate-level utilities corresponding the indirect, individual-level utility structure and the estimated parameters. These direct, aggregate-level utilities are the same regardless of the resulting nesting structure; the nesting structure influences the direct, aggregate-level utilities only through the estimated parameter values. More detailed discussion on parameter estimation is provided below.

Utility theory dictates that if individual-level utility functions are in the *Gorman form* they can be restated at the aggregate level by replacing individual-level variables with the corresponding aggregate-level variables for a given cohort (e.g., an origin zone). The Gorman form requires the individual-level utility function to consist of two separable components, specifically, an individual-specific component and a component that represents a common response to costs given a specified budget level (Varian 1992). These restrictions are mild for travel demand utilities, particularly since the budget constraint can be dropped from the problem (Oppenheim 1995).

Utility functions such as nested logit-based functions commonly represent *choice-specific* utilities, i.e., the utility obtainable from a particular decision, *given that choice*. However, solving for the utility-maximizing demands requires restating these indirect or *conditional* utilities as direct or *unconditional* utility functions that encapsulate preferences across *all* alternatives. This involves a straightforward mathematical transformation that does not impose any additional behavioral restrictions. Given an indirect utility structure, we can formulate a constrained optimization problem that maximizes indirect utility subject to a budget constraint. Then, finding the *dual* or mirror of the indirect utility maximization problem provides the corresponding direct utility function (Varian 1992).

The optimization problems that generate the equilibrium aggregate travel demands follow directly after deriving the aggregate-level direct utilities. The objective function of the optimization problems is a simple additive function of the direct utilities, i.e., including a travel demand component simply requires adding that direct utility component to the objective function. Similarly, it is simple matter to append the constraint set with the additional constraint required for the travel demand component. For example, given the direct utilities and constraints in Appendix 14.5.1 (14-89)-(14-101), solving for mode and UO-S route demands only (i.e., TG and TD demands are given) involves the following optimization problem:

$$MIN \ \ (14\text{-}91) + (14\text{-}92)$$

subject to:

$$(14\text{-}96)$$
$$(14\text{-}97)$$

and the non-negativity constraints:

$$(14\text{-}98) - (14\text{-}101)$$

Solving for the complete travel demands (TG, TD, MS, NA) with UO-S/NA requires:

$$MIN \ \ (14\text{-}89) + (14\text{-}90) + (14\text{-}91) + (14\text{-}92)$$

subject to:

$$(14\text{-}94)$$
$$(14\text{-}95)$$
$$(14\text{-}96)$$
$$(14\text{-}97)$$

and the non-negativity constraints:

$$(14\text{-}98) - (14\text{-}101)$$

8.2.5.1.3 Data requirements and parameter estimation

Parameter estimation for the TC models occurs at either the individual or aggregate levels through standard procedures for estimating nested logit models (see Ben-Akiva and Lerman 1985). A difficulty occurs if the TC model includes a SUO NA component since choice utilities are endogenous to the model (i.e., they are part of the model solution since travel costs are a function of flow). This difficulty is shared with most stochastic network equilibrium formulations (Anas 1988). Oppenheim (1995) provides some guidelines for dealing with this difficulty. Also, as noted previously in this report, there have been some progress in estimating SUO NA parameters during the solution phase (Huang 1995). Continued research on this topic is required.

A nested utility structure such as the one in the TC is often interpreted as a temporal sequence, e.g., a MS/NA nesting could represent the decision sequence of first choosing a mode then a route within that mode. However, an equivalent interpretation is the nesting represents interrelationships among choices with *no* implication of a choice sequence. For example, a MS/NA nesting structure captures the shared modal attributes of route choices within a particular mode. These interrelationships and their effect on nesting are reflected in the estimated values of the logit parameters and consequent restrictions on those values (14-102), (14-103). In brief, there is an inverse relationship between the variance of an unobserved utility and its corresponding logit parameter. This implies a nesting structure where the variances decrease as we move from the top level to the bottom. In other words, the modeler must be more certain about the observed utility specification at the lower-levels of the nesting structure than the top levels. Often, there is good reason to suspect a certain nesting structure (i.e., TG/TD/MS/NA nesting structure reflects a reasonable expectation about decreasing randomness in the utility functions). If the parameter estimation process indicates a violation of these conditions, the nesting structure must be respecified to reflect the estimation results. Fortunately, this is easily accomplished in Oppenheim's (1995) model.

8.2.5.1.4 Solution procedures

The convex combinations and Evan's partial linearization algorithms can solve models derived within the TC framework. Convex combinations can solve the UO-S-NA case with congestion effects, while Evan's partial linearization algorithm can solve all other models with congestion effects. Other solution procedures are available if congestion effects are not considered, but these models are not discussed here since they are less relevant to urban travel demand analysis.

9. DISCUSSION

9.1 Model Summary

Table 9-1 provides a summary of the network equilibrium-based travel demand models reviewed in this report. The Sheffi (1985), T2 (Dial 1995b), Dafermos (1980), Janson (1991) and Fisk (1980) are strictly *network assignment* models (NA) predicting equilibrium flows on a congested network. Sheffi (1985) provides a static, deterministic equilibrium that assumes perfect rationality among travelers, no temporal fluctuations and no modal or link interactions (although the basic formulation can be extended to include this latter consideration). Janson (1991) extends this formulation to encompass temporal dynamics, albeit using discrete time intervals. T2 (Dial 1995b) relaxes the strict equilibrium implied in these models to encompass varying tradeoffs among cost function components (in particular, VOTs) among travelers. Dafermos (1980) provides a very general (albeit complex) NA model that accommodate interactions among modes and link flows.

The remaining models are combined travel demand models. They are "combined" in the sense that they provide a combined or simultaneous equilibrium of the travel demand components specified. Three models are also fixed in the sense that the analyst solves for the travel demand components stated in the model rather than specifying the travel demand components of interest. These are: i) the NA/TD model of Evans (1976); ii) the NA/MS/TD model of Florian and Nguyen (1978); and, iii) the STEM NA/MS/TD/TG model (Safwat and Magnanti 1988). Admittedly, labeling these models as fixed (inflexible) may be harsh: these models could conceivably be modified to account for different applications (e.g., the MS component of STEM can be easily removed by specifying a single mode network). However, these modifications are *ad hoc* rather than inherent model features.

The combined travel demand models by Dafermos (1982), Sheffi and Daganzo (1980) and Oppenheim (1995) allow more flexibility in tailoring the combined travel demand model to fit the application. The Dafermos (1982) model allows flexible specification of the MS, TD and TG components: these functions need only satisfy very general qualitative conditions. The super- and hypernetwork approach allows specification of any or all of the travel demand components: model specification occurs entirely by specifying the abstract network corresponding to desired travel demand components. Finally, the trip consumer approach of Oppenheim (1995) offers flexibility at several levels. Not only can the model accommodate any or all travel demand components but also allows detailed specification of the individual-level utility structure for travelers' decisions.

As the above summary indicates, a fairly wide range of combined travel demand models are available. Choosing a particular model can depend on a variety of factors, not the least of which are application-specific circumstances. The next subsection of this report provides a comparison among models to help guide this selection choice.

Model	NA	MS	TD	TG
Sheffi (1985)	Static Deterministic No modal/link interactions			
T2 (Dial 1995b)	Static Deterministic No modal/link interactions Varying value of time			
Dafermos (1980)	Static Deterministic Modal/link interactions			
Janson (1991)	Discrete-time, dynamic No modal/link interactions			
Fisk (1980)	Static Stochastic (logit) No modal/link interactions			
Evans (1976)	Static Deterministic No modal/link interactions		Doubly constrained spatial interaction model	
Florian and Nguyen (1978)	Static Deterministic No modal/link interactions	Two modes (one fixed costs, other varying costs) Binomial logit	Doubly constrained spatial interaction model	
STEM (Safwat and Magnanti 1988)	Static Deterministic No modal/link interactions	Simultaneous with route choice	Logit	Logit-based accessibility function
Dafermos (1982)	Static Deterministic Modal/link interactions	General	General	General
Super- and hyper-networks (Sheffi and Daganzo 1980)	Static Deterministic No modal/link interactions	Logit or probit	Logit or probit	Logit or probit
Trip consumer approach (Oppenheim 1995)	Static Deterministic or stochastic No modal/link interactions or two-mode, symmetric interactions	Nested logit	Nested logit	Nested logit

Table 9-1: Summary of equilibrium travel demand models

9.2 Model Comparison

This section compares the travel demand methods discussed previously in the report. This comparison provides guidance for model selection and use in forecasting and policy analysis. Note that a definitive answer will not be forthcoming in this section. Model selection depends on the analysis requirements, data availability, computational resources, and so on, that vary substantially among different analysts and organizations (and even among projects by the same analyst within the same organization). In addition, it is not necessary for an analyst to ✌buy into✍ a single model since data requirements and computational procedures can be shared among the different models. A GIS platform facilitates this integration and flexibility. These platform issues will be discussed subsequent to this section

Comparison of the travel demand models uses the following criteria (based partially on Dial 1995a). The first criterion is *basic theory*. This concerns the major strengths and weaknesses of the model’s theoretical base, i.e., how well does it represent accepted travel demand theory. The second criterion is *mathematical elegance*. This refers to the parsimony and flexibility of the model’s formalism. In this case, we are not concerned about the “correctness” of the model *per se* but rather its ability to adapt to different analysis needs in a straightforward manner. The third criterion is *computational requirements and performance*. This includes the basic procedural needs of each model’s algorithm as well as performance efficiency. The final criterion is *data requirements and parameter estimation*. This is also a very pragmatic concern: many practitioners may consider this to be the fundamental, “make or break” criterion. Table 9-2 provides a summary comparison based on these criteria. In this table, a “+” indicates a model strength and a “-“ indicates a model weakness.

Model	Basic theory	Mathematical elegance	Computational performance	Data and parameter estimation
Sheffi (1985)	UO-S (-) Separable cost functions (-)	Flexible link cost functions (+)	Convex combinations method (+)	Standard link cost function (+)
T2 (Dial 1995b)	UO-G (+) Separable cost functions (-) Explicit recognition of varying cost tradeoffs (+)	Flexible link cost functions (+)	T2-RSD algorithm with parametric tree-building algorithm (+)	No estimation theory or procedures specified (-)
Dafermos (1980)	UO-G (+) Non-separable cost functions (+)	Flexible link cost functions (+)	Current solution method restricted to small problems (-)	No estimation theory or procedures specified (-)
Janson (1991)	DUO (-/+) Separable cost functions (-)	Inflexible link cost functions (-)	DTA and CDA (+)	Standard link cost function (+)
Fisk (1980)	SUO (+) Separable cost functions (-)	Flexible link cost functions (+) UO-S can be derived as a special case (+)	Chen and Alfa (1991) modified MSA (+)	Difficult parameter estimation (-)
Evans (1976)	UO-S (-) Separable cost functions (-) Separable demand function (-)	Inflexible TD function (-)	Partial linearization algorithm (+)	No estimation theory or procedures specified (-)
Florian and Nguyen (1978)	UO-S (-) Separable demand functions (-)	Flexible link cost functions (+) Inflexible MS and TD functions (-)	Partial linearization-related solution algorithm (+)	Parameter estimation based on average trip length (+)
STEM (Safwat and Magnanti 1988)	UO-S (-) Separable cost functions (-) Separable demand functions (-)	Flexible link cost functions (+) Flexible TD and TG functions (+)	Convex combinations method (+)	No estimation theory or procedures specified (-)
Dafermos (1982)	UO-G (+) Non-separable cost functions (+) Non-separable demand functions (+)	Flexible link cost functions (+) Flexible MS, TD and TG components (+)	Current solution method restricted to small problems (-)	No estimation theory or procedures specified (-)
Super- and hyper-networks (Sheffi and Daganzo 1980)	UO-S for NA (-) Separable cost functions (-) Separable demand functions (-)	Flexible link cost functions (+) Flexible MS, TD and TG functions (+)	MSA slow to converge (-), although Chen and Alfa (1991) modification faster with logit route choice (+)	No estimation theory or procedures specified, although standard random utility theory estimation procedures are available (+)
Trip consumer approach (Oppenheim 1995)	UO-S (-) Separable cost functions (-) Separable demand functions (-) Nested logit links	Flexible link cost functions (+) Flexible MS, TD and TG functions (+) Other models can be derived as special cases (+)	Convex combinations method (+) Partial linearization algorithm (+)	No estimation theory or procedures specified, although nested logit foundation may provide basis for parameter estimation

	individual and aggregate travel demand (+)			(+)

Table 9-2: Comparison of equilibrium travel demand models.

9.2.1 *Basic theory*

Equilibrium travel demand models offer tradeoffs between theoretical soundness and computational tractability. To achieve computational tractability, most models rely on foundations such as UO-S network equilibrium and separable link cost functions. These require the analyst to accept behavioral assumptions such as perfect information and decision making in route choice, no modal or link flow interactions and a static equilibrium (Sheffi 1985; Evans 1976; Florian and Nguyen 1978; Safwat and Magnanti 1988; Sheffi and Daganzo 1980; Oppenheim 1995). Conversely, models with highly general and realistic equilibrium foundations such as UO-G with non-separable link cost functions suffer from computational difficulties (Dafermos 1980, 1982). Between these extremes are models that selectively relax the strict UO-S and separability requirements. These include allowing varying tradeoffs among cost components in route choice (Dial 1995b), extending UO-S to encompass temporal dynamics (Janson 1991) and allowing for imperfect decision-making and information (Fisk 1980).

Similarly, non-separable demand functions are more realistic than separable demand functions but impose computational difficulties (Dafermos 1982). Separable demand functions allow greater tractability but assume that mode and trip destination choice are based on travel costs independent of other destinations and modal costs. Travel demand models implement separable demand functions with varying degrees of theoretical adequacy. Several models use an individual-level logit-based random utility formulation or an equivalent, aggregate-level spatial interaction formulation for the MS, TD or TG components (Evans 1976; Florian and Nguyen 1978; Safwat and Magnanti 1988). The IIA properties inherent in these formulations cannot account for interdependencies among unobserved choice-dependent components; this is particularly problematic with MS since travel modes have substantial interrelationships. IIA properties can also be problematic with respect to destination choice since these ignore spatial structural effects, i.e., hierarchical decision-making related to perceived spatial clustering of destinations.

Two models provide a selective relaxation of the strict separable, logit or spatial interaction foundation for the higher-level travel demand components. The super- and hypernetwork approach can accommodate both logit-based or probit-based choice mechanisms for the higher-level travel demands. However, a probit-based choice mechanism involves some computational difficulties related to the need to numerically (as opposed to analytically) evaluate choice probabilities and slow convergence of the MSA algorithm (see the "Computational Performance" section below). The trip consumer approach (Oppenheim 1995) uses a nested logit approach to capture choice interdependencies without sacrificing tractability: nested logit choice probabilities can be evaluated analytically and efficient solution algorithms (convex combinations, the partial linearization algorithm) can solve the resulting combined model. However, neither approach captures the spatial structural effects in destination choice.

In summary, the network-equilibrium-based travel demand models offer tradeoffs between theoretical consistency and tractability. Generally, some theoretical correctness must be sacrificed in order to achieve computational tractability, particularly for urban-scale travel demand analyses. *Note that these theoretical restrictions are shared by the traditional four-step approach.* Continued research is required to reconcile the more general and theoretically-consistent approaches such as Dafermos (1980, 1982) with more efficient, tractable solution procedures. However, at present there are models that offer selective relaxation of the strict behavioral assumptions without sacrificing tractability, in particular the T2 (Dial 1995b) and DUO (Janson 1991) NA models and the trip consumer approach (Oppenheim 1995) for combined NA/MS/TD/TG.

9.2.2 *Mathematical elegance*

Travel demand analyses are often used in "what-if?" scenario evaluation and in projecting future travel demands. Both applications benefit from the ability to incorporate infrastrucure and policy variables, e.g., the analyst can manipulate these variables to assess the impact of proposed infrastucture and policy changes on travel demand patterns. This requires model flexibility with respect to link cost and travel demand functions: these functions should be able to incorporate a wide range of infrastructure and policy-related variables to be useful in planning and policy analysis.

Generally, the travel demand models reviewed are very flexible with respect to model specification. Link cost functions and higher-level travel demand utilities can be specified with arbitrary length and complexity as long as they obey very general qualitative restrictions (e.g., separability, non-negativity, increasing with flow). Of the travel demand models reviewed, only three are inflexible with respect to link cost or travel demand function specification. Janson's (1991) DUO model can only accommodate travel time in its link cost functions since cumulative route costs (travel times) are required in the model constraints. Evans (1976) and Florian and Nguyen (1978) models use a doubly constrained spatial interaction model for the TD component; this only allows travel costs to affect destination choice. In both cases the origin outflows and destination inflows are constrained to known totals, so this is not a severe restriction if the analyst has a current or projected O-D matrix.

The Florian and Nguyen (1978) model also has severe restrictions on the MS function specification: this component is restricted to two modes of which only one has flow-dependent travel costs. Another source of inflexibility in this model is the use of a single parameter to control mode and destination choice dispersion. This limits model fit to empirical modal split and destination choice patterns.

The trip consumer approach (Oppenheim 1995) has particular strengths with respect to mathematical elegance and flexibility. First, the same modeling framework can easily capture any or all of the travel demand components in an elegant and consistent manner: all travel demand utilities are stated in a theoretically-consistent manner through the nested utility structure. In addition, the TC approach can accommodate measured utilities of arbitrary length and form: this could allow any number of relevant policy variables to be encompassed. The TC approach also allows the calculation of rigorous economic measures such as the *consumer's surplus* (see Varian 1992) deriving from a given transportation policy. Finally, several of the other models (e.g., Sheffi 1985; Evans 1976; Safwat and Magnanti 1987) can be derived as special cases.

9.2.3 *Computational performance*

The computational effort required to solve each model is a very practical consideration, particularly for urban-scale applications. In general, the travel demand models reviewed in this report are computationally tractable for urban-scale travel demand analyses. This is not surprising since this was the major selection criterion for inclusion in this review. The only models that are not (currently) feasible for urban-scale analyses are the Dafermos (1980, 1982) formulations. Nevertheless, this review included these models due to their theoretic appeal and mathematical elegance. Continued research is required to determine more tractable solution algorithms for this general and flexible approach.

An interesting property of the solution algorithms for most of the models reviewed in the report is they are structurally equivalent at a deep level. The convex combinations, Evans partial linearization and MSA algorithms all share the following basic steps: i) *direction-finding* - given a current feasible solution, find an optimal direction (within the mathematical solution space) that improves the objective function; ii) *step-size* - given the improvement direction, how far should we move?; and, iii) *convergence test* - should we stop and accept the current solution based on the degree of change since the last solution or solutions? Even further, the direction-finding steps among the algorithms all share the same major computational requirement, that is, solving the set of shortest paths from each origin to all destinations based on the current flow costs. Beyond this basic computational step, the algorithms differ with respect to how flows are distributed among destinations based on these shortest paths.

In terms of complexity, most algorithms are dominated by the need to solve the shortest path trees from each origin during the direction-finding step of each iteration. Therefore, relative efficiency reduces to the question of how fast the algorithm converges. In this respect, algorithms whose direction-finding step are based on partial linearization (e.g., Evans partial linearization, Florin and Nguyen's modified Hitchcock algorithm, the LDT STEM algorithm) rather than full linearization (convex combinations) will converge faster due to wider adjustment of the O-D travel demands during each step. However, this relative advantage only exists with O-D matrices that have few non-zero elements (Boyce, LeBlanc and Chon 1988). Algorithms with step-size optimization (all of the above algorithms plus the Chen and Alfa (1991) modified MSA) will converge faster than algorithms with fixed step-sizes (MSA).

Three algorithms require an additional complexity dimension beyond calculating shortest path trees from each origin during each iteration. Dial's (1995b) T2-RSD algorithm for solving the T2 NA model requires solving, from each origin and within each iteration, a shortest path tree for each slope interval along the efficient frontier. However, a parametric tree-building algorithm that modifies the tree from the previous interval rather than rebuilding from scratch provides substantial computational savings. Janson's (1991b) DTA procedure requires solving a user-designated number of shortest path trees from each origin during each iteration. This number determines an incremental assignment of flows from each origin during each iteration. In addition, each iteration corresponds to a (small) time interval over the study period; this number can be large. CDA requires similar calculations. Therefore, although DTA and CDA are not intractable for urban-scale analyses, their run-times can be long. However, this weakness will become less problematical as the power of desktop computational platforms continues to increase at a geometric rate.

9.2.4 *Data and parameter estimation*

The data requirements for the equilibrium travel demand model are reasonable. As noted in the introduction to this report, in general these models require no additional data beyond the data required for the four-step approach. The only exception to this observation are the Dafermos (1980, 1982) models; these require estimation of mode/link interactions and travel disutility interaction matrices.

A weakness of many of the travel demand models reviewed in this report is the lack of a consistent parameter estimation theory and procedure. Since these models provide a consistent equilibrium among the travel demand components, the parameters associated with the travel demand components should be estimated in a simultaneous (as opposed to an ad-hoc, sequential) manner. However, none of the models offer a statistical theory for the combined distribution of the parameters; this is required for developing an efficient simultaneous estimation procedure. A major reason for the lack of this theory is that most of these models come from scientists with a operations research/engineering background rather than an econometric/statistical background. A window of opportunity exists for researchers with the proper statistical background to contribute greatly to this field.

Some of the models reviewed offer informal guidelines with respect to parameter estimation. Techniques are available for estimating the link cost function parameters in several NA models (e.g, Fisk 1991). Florian and Nguyen (1978) provide some suggestions for estimating their combined NA/MS/TD model, although the tractable estimation procedure requires a single parameter shared among the MS and TD components. Sheffi and Daganzo's (1980) super- and hypernetwork approach can use standard logit and probit estimation procedures, although still required are methodologies for simultaneous estimation among these components. The trip consumer approach (Oppenheim 1995) offers considerable promise for simultaneous parameter estimation since a unified and consistent utility structure (the nested logit structure) underlies the model. Oppenheim (1995) provides detailed discussion of parameter estimation *issues* but does not offer a consistent and efficient estimation procedure *per se*; a window of opportunity exists in this regard.

9.3 Continued Research and Development Issues

Although this report's objective is an accessible review of equilibrium travel demand models rather than the research frontiers, this review nevertheless suggests two major research and development issues. These concern requirements for wider application of equilibrium travel demand models.

9.3.1 *Research and development issue 1: Specification and development of a computational toolkit for equilibrium travel demand modeling.*

The increasing use of geographic information systems for transportation (GIS-T) provides an excellent platform for developing equilibrium travel demand modeling software. A major benefit of GIS-T is database management and decision support. In the former case, GIS provides an efficient platform for building and maintaining the transportation database. Although travel demand data requirements includes a full range of spatial, aspatial and network data, the common unifying characteristic is location-based referencing (Shaw 1993). A GIS can not only maintain the primary data using location references but can also calculate critical spatial and topological properties required in the travel demand model (e.g., joins between origin and destination centroids and the transportation network). In the latter case, a GIS provides query-support as well as mapping of model results within the geographic context of the study area and with other ancillary but supporting cartographic information. This benefits analyses and policy decision processes (Armstrong *et al.* 1992).

While database management and decision support are the core benefits of GIS-T, another benefit of a GIS is support for *modelbase management*. The user interface of a GIS facilitates the treatment of models and model components as encapsulated objects which hide procedural details from the user. Instead, users manipulate these entities based on their attributes, behavior and relationships with other entities. This allows users to concentrate on concepts and substantive issues (e.g., data requirements, solution properties) instead of implementation details (Khoshafian and Abnous 1995). Note that these benefits can be realized from *any* computational platform that supports user interfaces (and especially graphical user interfaces or GUIs); however, these capabilities enhance the core role of GIS as a database management and decision support technique for transportation modeling.

Using a modelbase management approach to developing equilibrium travel demand modeling software offers a second major benefit, that is, the ability to integrate different travel demand models based on shared computational requirements and model components. The detailed review of the equilibrium travel demand models in Section 8 and the discussion of the computational performance issue previously in this section clearly illustrates the common structure and components shared among the models. Note that several of the UO-S models are derived from the basic UO-S NA model simply by adding additional components to the objective function and corresponding constraints to the constraint set. In addition, the solution procedures share the same fundamental structure of "direction-finding, move-size, convergence test." Similarly, the UO-G models share the same fundamental variational inequality (VI) structure. In turn, a VI formulation can generalize the convex combinations method used in several models (Magnanti and Perakis 1993). Finally, the Fisk (1980) SUO NA model shares a similar structure to UO-S NA and can collapse to this latter model under certain parameter settings.

Identifying the fundamental objects in the equilibrium travel demand models can allow the specification of a *computational toolkit* to support *several* of the equilibrium travel demand models within the same computational platform. Since many key computational objects are shared, implementing multiple travel demand models within the same platform can be accomplished efficiently. Then, instead of forcing a travel demand analysis into the model available in a given GIS software, the practitioner can access the model or models most appropriate for the research question at hand. This could greatly improve the flexibility and relevance of equilibrium travel demand modeling.

Miller and Storm (1996) identified an effective generic GIS design to support equilibrium travel demand modeling. This design partitions the system's components among GIS and non-GIS platforms according to the components' functional requirements. The GIS serves as a spatial database manager and GUI to the modeling system. This includes a network database design that maximizes the likelihood of database integrity after updates. The design exploits the ability of a GIS to maintain route data structures with a one-to-many relationship with the underlying topological network. The GIS also provides a "scenario editor" and "result analyzer" that exploits its user interface, cartographic visualization and spatial query capabilities. The travel demand solution algorithm resides outside the GIS; this provides substantial computational savings since the computational overhead required to access data with user-built functions within a GIS can be high. Since equilibrium travel demand models use shortest path calculations and flow updating as their primary solution mechanisms, the GIS can transfer the information into a network data structure than can be supported and accessed independently. After achieving an equilibrium solution, the GIS can access this network-based data for visualization and query capabilities.

The heterogeneous design strategy discussed by Miller and Storm (1996) indicates the desirability of component sharing and interoperability among the objects in the computational toolkit. This would allow the travel demand models to be supported and interfaced across a variety of computational platforms and software. This requires the computational tookit components to have the following features (Khoshafian and Abnous 1995): i) *binary representation* or the ability of components to be written in different languages but have standard interfaces to support interoperability among objects; ii) *standard user interfaces* or a common, recognized way in which users interface with the object; iii) *standard storage representation* or a common method for storing components in a nested or hierarchical manner; and, iv) *distributed computing support* or standards for the interaction of components in a distributed architecture. These standards will allow interoperability among the travel demand toolkit components, computational platforms, GIS software and other, supporting software.

A crucial first step in developing an equilibrium travel demand modeling toolkit is a structured analysis of the model components in terms of their *signatures* (inputs and outputs) and *behaviors*. As the preceding discussion should suggest, an appropriate technique is *object-oriented analysis and design* (Booch 1994; Rumbaugh *et al*. 1991). This is a graphics-oriented formal modeling technique that specifies *what* a system is (analysis) and *how* it should be designed. Most importantly, this type of structured analysis technique embodies several of the critical features necessary for a robust and effective computational toolkit, including encapsulation of procedures for binary representation, identifying standard interfaces through abstract data types and standard storage representing using inheritance and aggregation. An object-oriented analysis and design will allow specification of standards for the equilibrium travel demand toolkit components; this will allow GIS or other software vendors to develop interoperable and reusable components at the onset rather than having to re-engineer these at a later date.

9.3.2 *Research and development issue 2: Development of a model testbed*

Closely related to research and development issue 1 is the specification and development of a platform for extensive empirical testing of the equilibrium travel demand models. As noted previously, equilibrium travel demand models have only been subjected to limited testing (although these initial results are encouraging, at least with respect to the four-step approach). Continued and extensive testing and evaluation of the equilibrium travel demand models, as well as other competing approaches, is warranted.

The basic idea is to develop a model *testbed* that will support empirical evaluation of a variety of travel demand models. The "testbed" should be a robust and flexible computational platform that will support goodness-of-fit comparisons among different travel demand models. This should include modules that support the travel demand models, simulation of travel demand scenarios, graphics for summarizing model fit and a software development environment for generating required software components (see Summers and Southworth 1998). This testbed should not be restricted to equilibrium travel demand models; other approaches such as microsimulation-based models should be supported and tested within the system.

Development of a travel demand modeling testbed will be imperative given the continued development and deployment of intelligent transportation systems (ITS). While theory can not be ignored, identifying ITS-appropriate models cannot be identified only from first principles. The close coupling of transportation systems with ITS dictates the need for model validity *relative to the type of control imposed by the ITS* in addition to the traditional validity relative to an empirical dataset. Since varied ITS environments will dictate diverse modeling approaches, the testbed should support the economical development and testing of travel demand models relative to planned ITS deployments (Summers and Southworth 1998). The continuing development of dynamic equilibrium travel demand models can support ITS, although extensive testing of these and other approaches is required.

9.3.3 *Research and development issue 3: Development of a combined statistical distribution theory and simultaneous parameter estimation procedures.*

As mentioned previously in this section, a weakness of equilibrium travel demand models is a lack of statistical distribution theory for the combined travel demand components within each equilibrium model. Note that this weakness is shared with the 4-step approach: a consistent combined statistical distribution theory does not exist for the sequential travel demand estimation procedure. However, this weakness is not as apparent in the 4-step approach since it artificially separates the travel demand modeling components. When these components are embedded in an equilibrium framework, this weakness becomes more obvious.

Some discussion of these combined estimation issues does exist in the literature (e.g., Anas 1988; Florien and Nguyen 1978; Oppenheim 1995). However, no existing model has a combined statistical distribution theory and an efficient and unbiased simultaneous estimation procedure for all parameters. A possible theoretical foundation for this theory and procedure may be derived using the entropy-maximizing framework of Wilson (1967, 1974). This framework demonstrates that spatial interaction models provide the most likely trip distribution given known aggregate information about the system (e.g., origins outflows, destination inflows, travel costs, total trips). This theory supports a highly general information-minimizing approach to spatial interaction parameter estimation (see Fotheringham and O'Kelly 1989). The difficulty in combined travel demand parameter estimation is that travel costs, a key component of the system, are endogenous to the model (see Anas 1988 for a related discussion).

Oppenheim's (1995) trip consumer (TC) model provides possibly the best model support for parameter estimation. The TC model is consistent with a combined, nested logit utility structure at the individual level. The nested logit structure can support effective simultaneous parameter estimation. Oppenheim (1995, Chp. 7) provide a lucid discussion of the parameter estimation issues, particularly with respect to maximum likelihood estimation procedures. Continued research along this lines, most likely using the TC model as a basis, is required for effective application of equilibrium travel demand models.

10. CONCLUSION

A potentially accessible modeling framework that resolves the major flaws of the 4-step approach exists in the travel demand analysis literature. Equilibrium travel demand models generate consistent estimates of trip generation, trip distribution, modal split and network assignment without major increases in computational nor data requirements relative to the four-step approach. In addition, recent research has improved the behavioral generality of these models, linked aggregate travel demands to individual-level choice theory in a theoretically-consistent manner and developed linkages to dynamic travel demand estimation.

This research report provides a guide to the theory and practice of equilibrium travel demand modeling. The orientation of this report towards an accessible review intends to disseminate this information among a wide audience of current and emerging transportation planners and analysts. In addition, this report briefly identifies two major research and development issues to support wider application of equilibrium travel demand models.

As Boyce, Zhang and Lupa (1994) argue, continued progress in improving travel demand forecasts can only occur with increased understanding of the equilibrium approach. Professionals must insist that vendors provide software that implement the equilibrium approach and instructors must train the emerging generation of transportation planners and analysts in these modeling principles. This report supports this view by providing an initial step towards wider application of the equilibrium approach.

11. LITERATURE CITED

Aashtiani, H. Z., Magnanti, T. L. (1981) "Equilibria on a congested transportation network," *SIAM Journal of Algebraic and Discrete Methods*, 2, 213-226.

Anas, A. (1988) "Statistical properties of mathematical programming models of stochastic network equilibrium," *Journal of Regional Science*, 28, 511-530.

Armstrong, M. P., Densham, P. J., Lolonis, P. and Rushton, G. (1992) "Cartographic displays to support locational decision making," *Cartography and Geographic Information Systems*, 19, 154-164.

Barrett, C., Berkbigler, K., Smith, L., Loose, V., Beckman, R., Davis, J., Roberts, D. and Williams, D. (1995) "An operational description of TRANSIMS," Report LA-UR-95-2393, Los Alamos National Laboratory.

Batty, M. and Mackie, S. (1972) "The calibration of gravity, entropy and related models of spatial interaction," *Environment and Planning A*, 4, 205-233.

Beckmann, M., McGuire, G. B. and Winsten, C. B. (1956) *Studies in the Economics of Transportation*, New Haven, CT: Yale University Press.

Bell, M. G. H. (1991) "The estimation of origin-destination matrices by constrained generalised least squares," *Transportation Research B*, 25B, 13-22.

Ben-Akiva, M., Bolduc, D. and Bradley, M. (1993) "Estimation of travel choice models with randomly distributed values of times," *Transportation Research Record*, 1413, 88-97.

Ben-Akiva, M. and Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press.

Booch, G. (1994) *Object-oriented Analysis and Design with Applications*, Redwood City, CA: Benjamin/Cummings.

Boyce, D. E. (1984) "Urban transportation network-equilibrium and design models: Recent achievements and future prospects," *Environment and Planning A*, 16, 1445-1474.

Boyce, D. E., LeBlanc, L. J. and Chon, K. S. (1988) "Network equilibrium models of urban location and travel choices: A retrospective survey," *Journal of Regional Science*, 28, 159-183.

Boyce, D. E., Zhang, Y.-F. and Lupa, M. R. (1994) “Introducing feedback into four-step travel forecasting procedure versus equilibrium solution of combined model,” *Transportation Research Record* , 1443, 65-74.

Branston, D. (1976) “Link capacity functions: A review,” *Transportation Research*, 10, 223-236.

Cantarella, G. E. and Cascetta, E. (1995) “Dynamic processes and equilibrium in transportation networks: Towards an unifying theory,” *Transportation Science*, 29, 305-329.

Cascetta, E. and Nguyen, S. (1988) “A unified framework for estimating or updating origin/destination matrices from traffic counts,” *Transportation Research B*, 22B, 437-455.

Chen, M. and Alfa, A. S. (1991) “Algorithms for solving Fisk's stochastic traffic assignment model,” *Transportation Research B*, 25B, 405-412.

COMSIS (1996) *Incorporating Feedback in Travel Forecasting: Methods, Pitfalls and Common Concerns*, Final Report to Federal Highway Administration, U.S. Department of Transportation, Contract number DTFH61-93-C-00216, COMSIS Corporation.

Dafermos, S. (1972) “The traffic assignment problem for multiclass-user transportation networks,” *Transportation Science*, 6, 73-87.

Dafermos, S. (1980) “Traffic equilibrium and variational inequalities,” *Transportation Science*, 14, 42-54.

Dafermos, S. (1982) “The general multimodal network equilibrium problem with elastic demand,” *Networks*, 12, 57-72.

Daganzo, C. F. (1982) “Unconstrained extremal formulation of some transportation equilibrium problems,” *Transportation Science*, 16, 332-360.

Daganzo, C. F. and Sheffi, Y. (1977) “On stochastic models of traffic assignment,” *Transportation Science*, 11, 253-274.

Damberg, O., Lundgren, J. T. and Patriksson, M. (1996) “An algorithm for the stochastic user equilibrium problem,” *Transportation Research B*, 30, 115-131.

Dial, R. B. (1971) “A probabilistic multipath traffic assignment algorithm which obviates path enumeration,” *Transportation Research*, 5, 83-111.

Dial (1995a) "Network-optimized congestion pricing: A parable, model algorithm and heuristic," manuscript, Volpe National Transportation Systems Center, Cambridge, MA; submitted to *Operations Research*.

Dial (1995b) "T2: Another multipath probabilistic traffic assignment model that obviates path enumeration," manuscript, Volpe National Transportation Systems Center, Cambridge, MA; submitted to *Transportation Research B*.

Dial, R. B. (1996) "Bicriterion traffic assignment: Basic theory and elementary algorithms," *Transportation Science*, 30, 93 - 111.

Evans, S. P. (1976) "Derivation and analysis of some models for combining trip distribution and assignment," *Transportation Research*, 10, 37-57.

Fernandez, J. E. and Friesz, T. L. (1983) "Equilibrium predictions in transportation markets: The state of the art," *Transportation Research B*, 17B, 155-172

Fisk, C. S. (1980) "Some developments in equilibrium traffic assignment," *Transportation Research B*, 14B, 243-255.

Fisk, C. S. (1991) "Link travel time functions for traffic assignment," *Transportation Research B*, 25B, 103-113

Fisk, C. S. and Boyce, D. E. (1983) "A note on trip matrix estimation from traffic count data," *Transportation Research B*, 23B, 245-250.

Florian, M. and Nguyen, S.(1976) "An application and validation of equilibrium trip assignment methods," *Transportation Science*, 10, 374-390.

Florian, M. and Nguyen, S. (1978) "A combined trip distribution, modal split and trip assignment model," *Transportation Research B*, 12, 241-246.

Florian, M., Nguyen, S. and Ferland, J. (1975) "On the combined distribution-assignment of traffic," *Transportation Science*, 9, 43-53.

Fotheringham, A. S. and O'Kelly, M. E. (1989) *Spatial Interaction Models: Formulations and Applications*, Dordrecht: Kluwer Academic.

Friesz, T. L. (1985) "Transportation network equilibrium, design and aggregation: Key developments and research opportunities," *Transportation Research A*, 19A, 413-427.

Friesz, T. L., Bernstein, D., Mehta, N. J., Tobin, R. L. and Ganjalizadeh, S. (1994) "Day-to-day dynamic network disequilibria and idealized traveler information systems," *Operations Research*, 42, 1120-1136.

Friesz, T. L., Bernstein, D. and Stough, R. (1996) "Dynamic systems, variational inequalities and control theoretic models for predicting time-varying urban network flows," *Transportation Science*, 30, 14-31.

Huang, H.-J. (1995) "A combined algorithm for solving and calibrating the stochastic traffic assignment problem," *Journal of the Operational Research Society*, 46, 977-987.

Janson, B. N. (1991a) "A convergent algorithm for dynamic traffic assignment," *Transportation Research Record*, 1328, 69-80.

Janson, B. N. (1991b) "Dynamic traffic assignment for urban road networks," *Transportation Research B*, 25B, 143-161.

Janson, B. N. and Robles, J. (1993) "Dynamic traffic assignment with arrival time costs," in C. F. Daganzo (ed.) *Transportation and Traffic Theory*, Amsterdam: Elsevier Science, 127-146.

Janson, B. N. and Robles, J. (1995) "Quasi-continuous dynamic traffic assignment model," *Transportation Research Record*, 1493, 199-206.

Janson, B. N. and Southworth, F. (1992) "Estimating departure times from traffic counts using dynamic assignment," *Transportation Research B*, 26B, 3-16.

Jayakrishnan, R., Tsai, W. K., Prashker, J. N. and Rajadhyaksha, S. (1994) "Faster path-based algorithm for traffic assignment," *Transportation Research Record*, 1443, 75-83.

Khoshafian, S. and Abnous, R. (1995) *Object Orientation*, New York: John Wiley and Sons.

Kulkarni, R. G., Stough, R. R., Haynes, K. E. (1996) "Spin glass and the interactions of congestion and emissions: An exploratory step," *Transportation Research C*, 4C, 407-424.

Lawphongpanich, S. and Hearn, D. W. (1984) "Simplicial decomposition of the asymmetric traffic assignment problem," *Transportation Research B*, 18B, 123-133.

Leurent, F. M. (1995) "Contributions to logit assignment model," *Transportation Research Record*, 1493, 207-212.

Magnanti, T. and Perakis, G. (1993) "A unifying geometric solution framework and complexity analysis for variational inequalities," Working Paper OR-276-93, MIT, Cambridge, MA; cited in Dial (1996).

Miller, H. J. and Storm, J. D. (1996) "Geographic information system design for network equilibrium-based travel demand models," *Transportation Research C*, 4, 373-389.

Nagurney, A. (1993) *Network Economics: A Variational Inequality Approach*, Dordrecht: Kluwer Academic.

Nguyen, S. (1984) "Estimating origin-destination matrices from observed flows," in M. Florian (ed.) *Transportation Planning Models*, Amsterdam: Elsevier Science, 363-380.

Oppenheim, N. (1995) *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*, New York: John Wiley and Sons.

Ortuzar, J. de D. and Willumsen, L. G. (1990) *Modelling Transport*, New York: John Wiley and Sons.

Ran, B. and Boyce, D. E. (1994) *Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach*, Berlin: Springer.

Ran, B., Hall, R. W. and Boyce, D. E. (1996) "A link-based variational inequality model for dynamic departure time/route choice," *Transportation Research B*, 30. 31-46.

Rumbaugh, J., Blaha, M., Prenerlani, W., Eddy, F. and Lorensen, E. (1991) *Object-oriented Modeling and Design*, Englewood Cliffs, N. J.: Prentice-Hall.

Safwat, K. N. A. (1987a) "Application of a simultaneous transportation equilibrium model to intercity passenger traffic in Egypt," *Transportation Research Record*, 1120, 52-59.

Safwat, K. N. A. (1987b) "Computational experience with a convergent algorithm for the simultaneous prediction of transportation equilibrium," *Transportation Research Record*, 1120, 60-67.

Safwat, K. N. A. and Magnanti, T. L. (1988) "A combined trip generation, trip distribution, modal split, and trip assignment model," *Transportation Science*, 18, 14-30.

Safwat, K. N. A. and Walton, C. M. (1988) "Computational experience with an application of a simultaneous transportation equilibrium model to urban travel in Austin, Texas," *Transportation Research B*, 22B, 457-467.

Sen, A, and Smith, T. E. (1995) *Gravity Models of Spatial Interaction Behavior*, Berlin: Springer-Verlag.

Shaw, S.-L. (1993) "GIS for urban travel demand analysis: Requirements and alternatives," *Computers, Environment and Urban Systems*, 17, 15-29.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Englewood Cliffs, N.J.: Prentice-Hall.

Sheffi, Y. and Daganzo, C. F. (1978) "Hypernetworks and supply-demand equilibrium obtained with disaggregate demand models," *Transportation Research Record*, 673, 113-121.

Sheffi, Y. and Daganzo, C. F. (1980) "Computation of equilibrium over transportation networks: the case of disaggregate demand models," *Transportation Science*, 14, 155-173.

Sheffi, Y. and Powell, W. B. (1982) "An algorithm for the equilibrium assignment problem with random link times," *Networks*, 12, 191-207.

Slavin, H. (1995) "An integrated, dynamic approach to travel demand forecasting," *Transportation*, 22, 1-40.

Smith, M. J. (1979) "The existence, uniqueness and stability of traffic equilibria," *Transportation Research B*, 13B, 295-304.

Southworth, F. (1995) "A technical review of urban land use-transportation models as tools for evaluating vehicle travel reduction strategies," research report ORNL-6881, Center for Transportation Analysis, Energy Division, Oak Ridge National Laboratory; available at www.bts.gov.

Summers, M. and Southworth, F. (1998) "Design of a testbed to assess alternative traveler behavioral models within an intelligent transportation system architecture," *Geographical Systems*, forthcoming.

Speiss, H. (1987) "A maximum likelihood model for estimating origin-destination matrices," *Transportation Research B*, 21B, 395-412.

Varian, H. R. (1992) *Microeconomic Analysis*, 3ed, New York: W.W. Norton.

Von Hohenbalken, B. (1977) "Simplicial decomposition in nonlinear programming algorithms," *Mathematical Programming*, 13, 49-68.

Wardrop, J. G. (1952) “Some theoretical aspects of road traffic research,” *Proceedings of the Institute of Civil Engineers, Part II*, 1(36), 325-362.

Webber, M. J. (1977) “Pedagogy again: What is entropy?” *Annals of the Association of American Geographers*, 67, 254-266.

Williams, P. A. and Fotheringham, A. S. (1984) *The Calibration of Spatial Interaction Models by Maximum Likelihood Estimation with Program SIMODEL*, Geographic Monograph Series 7, Department of Geography, Indiana University.

Wilson, A. G. (1967) “Statistical theory of spatial trip distribution models,” *Transportation Research*, 1, 253-269.

Wilson, A. G. (1974) *Urban and Regional Models in Geography and Planning*, London: John Wiley and Sons.

Wrigley, N. (1985) *Categorical Data Analysis for Geographers and Environmental Scientists*, London: Longman.

Yang, H., Iida, Y. and Sasaki, T. (1991) “An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts,” *Transportation Research B*, 25B, 351-363.

Yang, H., Iida, Y. and Sasaki, T. (1994) “The equilibrium origin-destination matrix estimation problem,” *Transportation Research B*, 28B, 23-33.

12. APPENDIX: SUMMARY OF MAJOR NOTATION AND DEFINITIONS

This appendix summarizes the basic notation and network flow properties that characterize network equilibrium-based travel demand models. The notation follows (but expands on) Fernandez and Friesz (1983).

12.1 Basic Notation

Network		
$G = [N, A]$	Directed graph representing transportation network, where N is a finite set of network nodes and A is a set of network arcs	**(12-1)**
I	Set of origins, $I \subseteq N$	**(12-2)**
J	Set of destinations, $J \subseteq N$	**(12-3)**
a	A network arc $a \equiv (n_l, n_m); n_l, n_m \in N$	**(12-4)**
r	A network path $r \equiv \{(n_i, n_k), (n_k, n_l), \ldots, (n_l, n_m), (n_m, n_j)\}$	**(12-5)**
R	Set of all paths in G	**(12-6)**
R_{ij}	Set of all paths that connect O-D pair i,j	**(12-7)**
K	Set of modes	**(12-8)**
d_{ar}^{k}	Arc-path incidence variable; equal to one if arc a belongs to path r and allows flows by mode k	**(12-9)**

Arc flows and costs		
f_a^k	Mode k flow on arc a	**(12-10)**
f_a	Total flow on arc a	**(12-11)**
F	Set of all arc flows	**(12-12)**

c_a^k	Average travel cost for mode k user on arc a	**(12-13)**
p_a^k	Random variable representing perceived travel cost for mode k user on arc a	**(12-14)**

Path flows and costs		
h_r^k	Mode k flow on path r	**(12-15)**
H	Set of all path flows	**(12-16)**
C_r^k	Average travel cost for mode k user on path r; $C_r^k = \sum_{a \in A} d_{ar}^k c_a^k$	**(12-17)**
C_{ij*}^k	Minimum average travel cost for mode k user between O-D pair *ij*	**(12-18)**
C_*	Set of minimum travel costs for all modes and O-D pairs	**(12-19)**
P_r^k	Random variable representing the perceived travel cost for mode k on path r	**(12-20)**
M_r^k	Marginal travel cost on path r for a mode k user; $M_r^k = \frac{\P C_r^k}{\P h_r^k}$	**(12-21)**
M_{ij*}^k	Minimum marginal travel cost between O-D pair *ij* for mode k user	**(12-22)**

Aggregate travel demands		
D_{ij}^k	Total mode k flow between O-D pair *i,j*	**(12-23)**
$d_{ij}^k \equiv \left(D_{ij}^k\right)^{-1}$	*Travel disutility* associated with mode k travel between O-D pair *i,j*	**(12-24)**
D_{ijr}	Total flow on path r between O-D pair *i,j*	**(12-25)**
D_i	Total outflow from origin i	**(12-26)**
D_j	Total inflow to destination j	**(12-27)**
D_{i0}	Non-travelers in origin i	**(12-28)**

B_i	Number of potential travelers in origin i	**(12-29)**

12.2 Special case notation

Dynamic flow notation (Janson 1991b)		
t	Discrete time interval	**(12-30)**
d	Origin departure time interval	**(12-31)**
Δt	Length of each time interval	**(12-32)**
T	Total number of time intervals	**(12-33)**
h_r^d	Flow on path r that departed during time interval d	**(12-34)**
C_r^d	Average travel cost on path r for travelers who departed during time interval d	**(12-35)**
C_{ij*}^d	Minimum average travel cost between O-D pair i,j for travelers who departed during time interval d	**(12-36)**
d_{ra}^d	Temporal arc-path incidence variable; equal to one if trips departing during time interval d and assigned path r use arc a during time interval t, zero otherwise	**(12-37)**
b_{rn}^d	Travel time of path r from its origin to node n for travelers departing in time interval d	**(12-38)**
A_n	Set of all arcs incident to node n	**(12-39)**

T2 notation (Dial 1995a, 1995b, 1996)		
$d_a(f_a)$	a deterministic disutility (*d-disutility)* associated with flow on arc a	**(12-40)**
$s_a(f_a)$	a stochastically-weighted disutility (*s-disutility)* associated with flow on arc a	**(12-41)**
w	a stochastic parameter (*s-weight*) capturing varying reactions among travelers to $s_a(f_a)$	**(12-42)**

$g(w\|i,j)$	s-weight's fixed and known probability density function specific to O-D pair *i*,*j*	**(12-43)**
$f_{ia}(w)$	Flow from origin *i* with s-weight ❄ on arc *a*	**(12-44)**
f_{ia}	Total flow on arc *a* from origin *i*	**(12-45)**
$\{a: a=(n_k, j)\}$	set of arcs whose to-nodes are destinations	**(12-46)**
$\{a: a=(j, n_k)\}$	set of arcs whose from-nodes are destinations	**(12-47)**

Super- and hyper-network notation (Sheffi and Dagnazo 1980)		
N	Set of basic nodes	**(12-48)**
V	Set of non-basic or "virtual" nodes, $I, J \subseteq V$	**(12-49)**
A	Set of basic arcs $A \equiv \{(n_l, n_m): l, m \in N\}$	**(12-50)**
E	Set of entrance/egress arcs $E = \{(n_i, n_l), \ldots, (n_m, n_j): i \in V, l \in N_i', m \in N_j'', j \in V\}$	**(12-51)**
N_i'	Set of basic nodes connected to origin *i* (i.e., "outbound" basic nodes connected to *i*)	**(12-52)**
N_j''	Set of basic nodes connected to destination *j* (i.e., "inbound" basic nodes connected to destination *j*)	**(12-53)**
N'	Set of all outbound basic nodes	**(12-54)**
N''	Set of all inbound basic nodes	**(12-55)**

12.3 Additional Definitions

12.3.1 *Basic flow feasibility requirements*

$$\{H\} \geq 0 \qquad \textbf{(12-56)}$$

$$\sum_{r\in R_{ij}} h_r^k = D_{ij}^k \quad \forall\,(i,j,k) \tag{12-57}$$

$$f_a^{\,k} = \sum_{r\in R} d_{\,ar}^{\,k}\, h_r^k \quad \forall\,(a,k) \tag{12-58}$$

12.3.2 *Separable versus non-separable cost functions*

Separable cost functions:

$$c_a^k = c_a^k(\, f_a^{\,k}) \tag{12-59}$$

Non-separable cost functions:

$$c_a^k = c_a^k(F) \tag{12-60}$$

12.3.3 *Separable versus non-separable demand functions*

Separable demand functions:

$$D_{ij}^k = D_{ij}^k\,(C_{ij^*}^k) \tag{12-61}$$

Non-separable demand functions:

$$D_{ij}^k = D_{ij}^k(C_*) \tag{12-62}$$

12.3.4 *Other conditions*

Cost function non-negativity:

$$x \ge 0 \Rightarrow c_a(x) \ge 0 \tag{12-63}$$

Cost function increasing with respect to flow levels:

$$\frac{\partial\, c_a(x)}{\partial\, x} > 0 \quad \forall\, a \in A \tag{12-64}$$

13. APPENDIX: FORMAL PROPERTIES OF TRANSPORTATION EQUILIBRIA

13.1 Network Equilibria

13.1.1 *User optimal (UO)*

13.1.1.1 User optimal - strict conditions (UO-S)

A vector of path flows H is a UO-S flow if it is *feasible* and (Fernandez and Friesz 1983):

$$h_r^k > 0 \Rightarrow C_r^k = C_{ij^*}^k \quad \forall\,(i, j, k, r \in R_{ij}) \tag{13-1}$$

$$C_r^k > C_{ij^*}^k \Rightarrow h_r^k = 0 \quad \forall\,(i, j, k, r \in R_{ij}) \tag{13-2}$$

13.1.1.2 User optimal - general conditions (UO-G)

A vector of path flows $\overline{H}$ is a UO-G flow if (Smith 1979):

$$C(\overline{H}) \cdot (H - \overline{H}) \geq 0 \quad \forall\; H \in \Omega \tag{13-3}$$

where Ω is the convex set of feasible path flows h_r^k.

13.1.2 *Dynamic user optimal (DUO)*

A vector of (discrete time) path flows H^d, $d = 1, \ldots, T$, is a dynamic user optimal flow if it is feasible (see equations (14-54)- (14-57)) and (Janson 1991b):

$$h_r^d > 0 \Rightarrow C_r^d = C_{ij*}^d \quad \forall\,(d \in T, r \in R_{ij}, i \in I, j \in J) \tag{13-4}$$

$$C_r^d \geq C_{ij*}^d \Rightarrow h_r^d = 0 \quad \forall\,(d \in T, r \in R_{ij}, i \in I, j \in J) \tag{13-5}$$

13.1.3 *System optimal (SO)*

A vector of path flows H is a SO flow if it is feasible and (Fernandez and Friesz 1983):

$$h_r^k > 0 \Rightarrow M_r^k = M_{ij^*}^k \quad \forall\, (i, j, k, r \in R_{ij}) \tag{13-6}$$

$$M_r^k > M_{ij^*}^k \Rightarrow h_r^k = 0 \quad \forall\, (i, j, k, r \in R_{ij}) \tag{13-7}$$

13.1.4 *Stochastic user optimal (SUO)*

A flow pattern H is a SUO if it is feasible and (Daganzo and Sheffi 1977; Sheffi 1985):

$$h_r^k = D_{ij}^k\, p_r^k \quad \forall \left(i, j, k, r \in R_{ij}\right) \tag{13-8}$$

where:

$$p_r^k = \Pr\left[P_r^k \le P_s^k \quad \forall\, r \ne s \in R_{ij} \mid \mathbf{C}\right] \tag{13-9}$$

$$P_r^k = c_r^k + e_r^k \tag{13-10}$$

$$\mathrm{E}\left[e_r^k\right] = 0 \Rightarrow \mathrm{E}\left[P_r^k\right] = c_r^k \tag{13-11}$$

where **C** is the vector of path travel costs.

13.2 Market Equilibrium

A flow pattern $\left(H, \mathbf{C}_*\right)$ is a UO-S-based market equilibrium with combined TG, TD, MS and NA if it satisfies the following system of non-linear equations (Aashtiani and Magnanti 1981; Fernandez and Friesz 1983):

$$(C_r^k(H) - C_{ij*}^k) h_r^k = 0 \quad \forall\, (i, j, k, r \in R_{ij}) \tag{13-12}$$

$$C_r^k(H) - C_{ij^*}^k \ge 0 \quad \forall\, (i, j, k, r \in R_{ij}) \tag{13-13}$$

$$\sum_{r \in R_{ij}} h_r^k - D_{ij}^k(C_*) = 0 \quad \forall\, (i, j, k) \tag{13-14}$$

$$C_r^k(H) = \sum_{a \in L} d_{ar}^{\,k} c_a^k(H) \quad \forall (i, j, k, r \in R_{ij}) \tag{13-15}$$

$$(H, C_*) \geq 0 \tag{13-16}$$

14. APPENDIX: MODEL FORMULATIONS

14.1 UO-S-based Approaches

14.1.1 *NA (Sheffi 1985)*

Assumptions:

i) one mode (although multi-mode extensions are possible;

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) D_{ij} fixed and exogenous.

Optimization problem:

$$\underset{\{f_a\}}{MIN} \quad \sum_a \int_0^{f_a} c_a(x)\, dx \qquad \textbf{(14-1)}$$

subject to:

$$f_a = \sum_{r \in R} d_{ar}\, h_r \quad \forall\, a \in A \qquad \textbf{(14-2)}$$

$$\sum_{r \in R_{ij}} h_r = D_{ij} \quad \forall\, i \in I, j \in J \qquad \textbf{(14-3)}$$

$$h_r \geq 0 \quad \forall\, r \in R \qquad \textbf{(14-4)}$$

14.1.2 *Combined TD/NA (Evans 1976)*

Assumptions:

i) one mode;

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) D_i, D_j fixed and exogenous;

vi) separable demand functions (12-61) with the following format:

$$D_{ij} \propto \exp\left(-b_{ij}\ C_{ij*}\right) \tag{14-5}$$

Optimization problem:

$$\underset{\{D_{ij},\, f_a\}}{MIN} \quad \frac{1}{b}\sum_i \sum_j \left(D_{ij} \ln D_{ij} - D_{ij}\right) + \sum_a \int_0^{f_a} c_a(x)\, dx \tag{14-6}$$

subject to:

$$\sum_{r \in R_{ij}} h_r = D_{ij} \quad \forall\, i \in I, j \in J \tag{14-7}$$

$$f_a = \sum_{r \in R} d_{ar}\, h_r \quad \forall\, a \in A \tag{14-8}$$

$$\sum_i D_{ij} = D_j \quad \forall\, j \in J \tag{14-9}$$

$$\sum_j D_{ij} = D_i \quad \forall\, i \in I \tag{14-10}$$

$$h_r \geq 0 \quad \forall\, r \in R \tag{14-11}$$

$$D_{ij} \geq 0 \quad \forall\, i \in I, j \in J \tag{14-12}$$

14.1.3 *Combined TD/MS/NA - Florian and Nguyen (1978)*

Assumptions:

i) two modes ("automobile" and "public transit");

ii) separable cost functions (automobile) (12-59);

iii) non-negative cost functions (automobile) (12-63);

iv) increasing cost functions (automobile) (12-64);

v) $c_a^{k_2}$ (public transit) fixed and exogenous;

vi) D_i, D_j fixed and exogenous;

vii) separable demand functions (12-61) with the following format:

$$D_{ij}^k \propto \exp\left(-b\ \ C_{ij*}^k\right) \quad \forall\, i \in I, j \in J, k = k_1, k_2 \tag{14-13}$$

viii) $$\frac{D_{ij}^{k}}{D_{ij}^{k_1}+D_{ij}^{k_2}}=\frac{\exp\left(b\ C_{ij*}^{k}\right)}{\exp\left(b\ C_{ij*}^{k_1}\right)+\exp\left(b\ C_{ij*}^{k_2}\right)}\quad \forall i\in I,\, j\in J, k=k_1,k_2 \tag{14-14}$$

where k_1 indicates automobile mode and k_2 indicates public transit.

Optimization problem:

$$\underset{\left\{D_{ij}^{k_1},D_{ij}^{k_2},f_a\right\}}{MIN}\quad b\sum_{i=1}^{I}\sum_{j=1}^{J}D_{ij}^{k_1}\ln D_{ij}^{k_1}+\sum_{i=1}^{I}\sum_{j=1}^{J}D_{ij}^{k_2}\left(b\ln D_{ij}^{k_2}+C_{ij*}^{k_2}\right)+\sum_{a\in A}\int_0^{f_a}c_a(x)dx \tag{14-15}$$

subject to:

$$\sum_{j=1}^{J}\left(D_{ij}^{k_1}+D_{ij}^{k_2}\right)=D_i\quad \forall i\in I \tag{14-16}$$

$$\sum_{i=1}^{I}\left(D_{ij}^{k_1}+D_{ij}^{k_2}\right)=D_j\quad j\in J \tag{14-17}$$

$$\sum_{r\in R_{ij}}h_r^{k_1}=D_{ij}^{k_1}\quad \forall i\in I, j\in J \tag{14-18}$$

$$f_a=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r\in R_{ij}}d_{ar}^{k_1}h_r^{k_1}+f_a^{k_2} \tag{14-19}$$

$$D_{ij}^{k}\geq 0\quad \forall\ i\in I, j\in J, k\in K \tag{14-20}$$

$$h_r^{k}\geq 0\quad \forall\ r\in R, k\in K \tag{14-21}$$

where $f_a^{k_2}$ is the public transit mode's contribution to flow on arc a; this may be zero if the public transit route is separate from the street network.

14.1.4 *Combined TG/TD/MS/NA - STEM (Safwat and Magnanti 1988)*

Assumptions:

i) $$G'=\left[N',A'\right],\ N'=\bigcup_{k=1}^{K}N^k,\ A'=\bigcup_{k=1}^{K}A^k \tag{14-22}$$

where N^k,A^k are the nodes and arcs of the network for mode k;

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) separable demand functions (12-61) with the following format:

$$D_{ij} = D_i \frac{\exp\left(-q\, C_{ij*} + B_j\right)}{\sum_{k\in I} \exp\left(-q\, C_{ik*} + B_k\right)} \qquad \textbf{(14-23)}$$

vi) $D_i = a_i S_i + E_i \;\forall\, i \in I$ **(14-24)**

where:

$S_i =$ $$\max\left\{0, \ln \sum_{j\in J} \exp\left(-qC_{ij*} + E_j\right)\right\} \qquad \textbf{(14-25)}$$

accessibility variable that measures the expected utility of travel from origin i (endogenous);

$E_i =$ composite variable measuring the effect of non-transportation factors on travel flow from origin i (exogenous); **(14-26)**

$B_j =$ composite variable measuring the attractiveness of destination j (exogenous); **(14-27)**

Optimization problem ($K \equiv 1$ since modal attributes are captured by the subnetworks):

$$\underset{\{S_i, D_{ij}, f_a\}}{MIN} \quad \sum_i \frac{1}{q_i}\left[a_i S_i^2 + a_i S_i - \left(a_i S_i + E_i\right)\ln\left(a_i S_i + E_i\right)\right] + \sum_i \frac{1}{q_i} \sum_j \left(D_{ij} \ln D_{ij} - B_j D_{ij} - D_{ij}\right) + \sum_{a\in A'} \int_0^{f_a} c_a(x)dx \qquad \textbf{(14-28)}$$

subject to:

$$\sum_{j\in J} D_{ij} = a_i S_i + E_i \;\forall\, i \in I \qquad \textbf{(14-29)}$$

$$\sum_{r\in R_{ij}} h_r = D_{ij} \quad \forall\, i \in I, j \in J \qquad \textbf{(14-30)}$$

$$f_a = \sum_{r\in R} d_{ar}\, h_r \quad \forall\, a \in A' \qquad \textbf{(14-31)}$$

$$S_i \geq 0 \;\forall\, i \in I \qquad \textbf{(14-32)}$$

$$D_{ij} \geq 0 \quad \forall\, i \in I, j \in J \qquad \textbf{(14-33)}$$

$$h_r \geq 0 \quad \forall\, r \in R$$ **(14-34)**

14.2 UO-G-based approaches

14.2.1 *T2 NA (Dial 1995b)*

Assumptions:

i) one mode;

ii) separable cost functions (12-59) in the following format:

$$c_a \equiv d_a\left(f_a\right) + w\ s_a\left(f_a\right)$$ **(14-35)**

iii) D_{ij} fixed and exogenous

Optimization problem:

A flow pattern $F^* = \left\{ f_{ia}^* \right\}$ is user optimal-T2 (UO-T2) if and only if it is feasible and a solution to the following variational inequality problem:

$$\int_\Omega \sum_{i \in I} \sum_{a \in A} \left(d_a\left(f_a^*\right) + w\ s_a\left(f_a^*\right) \right) \left(f_{ia}(w) - f_{ia}^*(w) \right) dw \quad \geq 0$$ **(14-36)**

subject to:

$$\sum_{\{a:a=(x_k,j)\}} f_{ia}(w) - \sum_{\{a:a=(j,x_k)\}} f_{ia}(w) = D_{ij}\ g\left(w \middle| i, j\right) \quad \forall\, j \in J$$ **(14-37)**

14.2.2 *NA/MS (Dafermos 1980)*

Assumptions:

i) one or more modes;

ii) non-separable cost functions (12-60) in the following format:

$$c_a(F) = \mathbf{G} \cdot F + \mathbf{b}$$ **(14-38)**

where:

G =	a matrix capturing the interactions among links in the network	**(14-39)**
b =	a vector containing static (e.g., base) costs for each network arc.	**(14-40)**

iii) Jacobian matrix of the cost functions, $\left[\frac{\partial c_a^k}{\partial F}\right]$, is positive definite **(14-41)**

Optimization problem:

A flow pattern $\overline{F}$ is UO-G if and only if it feasible ((14-2) - (14-4)) and:

$$\mathbf{c}\left(\overline{F}\right)\left(F-\overline{F}\right)\geq 0 \quad \forall \quad F\in k \tag{14-42}$$

where $\mathbf{c}$ is the vector of all arc costs and k is the set of feasible arc flows.

14.2.3 *Combined TG/TD/MS/NA (Dafermos 1982)*

i) one or more modes;

ii) non-separable cost functions (12-60) in the format of (14-38);

iii) Jacobian matrix of the cost functions, is positive definite (14-41);

iv) non-separable demand functions (12-62) in the following format

$$d_{ij}^k\left(D_{ij}^k\right)=\mathbf{M}\cdot D_{ij}^k+\mathbf{s} \tag{14-43}$$

where:

$\mathbf{M}=$	a matrix providing travel disutility interactions among O-D flows.	**(14-44)**
$\mathbf{s}=$	a vector containing static (e.g., base) disutilities between O-D pairs.	**(14-45)**

v) Jacobian matrix of the inverse demand functions, $\left[\frac{\partial d_{ij}^k}{\partial D_{ij}^k}\right]$, is positive definite **(14-46)**

Given the assumptions above, an (arc) flow and travel demand pattern $(\overline{F},\overline{D})$ is a market equilibrium with combined TG/TD/MS/NA if it satisfies the following variational inequality problem (Dafermos 1982):

$$\mathbf{c}(\overline{F})(F-\overline{F})-\mathbf{d}(\overline{D})(D-\overline{D})\geq 0 \quad \forall\, F,D\in\Gamma \qquad \textbf{(14-47)}$$

where **d** is the vector of all travel disutilities (12-24) and Γ is the set of feasible flow patterns and travel demands.

The VI problem (14-47) is a generalization of the following individual-level, UO-S-based market equilibrium conditions:

$$d_{ij}^k(D_{ij}^k)-C_r^k(F) \quad \begin{cases} =0, & if\ h_r^k>0 \\ \leq 0, & if\ h_r^k=0 \end{cases} \quad \forall\,(k,w,r\in R_w) \qquad \textbf{(14-48)}$$

If $(\overline{F},\overline{D})$ is a demand pattern that satisfies (14-48) then the following will be true:

$$C_r^k(\overline{F})(h_r^k-\overline{h}_r^k)-d_{ij}^k(\overline{D})(h_r^k-\overline{h}_r^k)\geq 0 \quad \forall\,(i,j,k,r\in R_{ij}) \qquad \textbf{(14-49)}$$

where $\overline{h}_r^k$ is the route flow implied by $\overline{F}$. Expressing the first half of (14-49) in terms of arc flows only and the second half in travel demands only and then summing across user classes and routes leads directly to (14-47). The aggregate level statement relaxes the strict UO-S assumptions and allows individual variations in behavior within the aggregate constraints.

14.3 DUO-based approaches

14.3.1 *DUO NA (Janson 1991b)*

Assumptions:

i) single mode;

ii) separable cost functions (12-61);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) D_{ij} fixed and exogenous

vi) study time period divided into discrete time intervals $t=1,\ldots,T$

Model structure:

$$\underset{\{f_a^t\}}{MIN} \quad \sum_{a\in A}\sum_{t\in T}\int_0^{f_a^t} c_a^t(x)\,dx \qquad \textbf{(14-50)}$$

subject to:

(static constraints)

$$f_a^t = \sum_{r \in R} \sum_{t_d \in T} h_r^d d_{ra}^{d\,t} \quad \forall\ a \in A, t \in T \tag{14-51}$$

$$D_{ij}^d = \sum_{r \in R_{ij}} h_r^d \quad \forall\ i \in I, j \in J, d\ \in T \tag{14-52}$$

$$h_p^d\ \geq 0 \quad \forall\ r \in R, d\ \in T \tag{14-53}$$

(dynamic contraints)

$$\sum_{t \in T} d_{ra}^{d\,t} = 1 \quad \forall r \in R, a \in A_p, t \in T, d\ \in T \tag{14-54}$$

$$b_{rn}^t = \sum_{t \in T} \sum_{a \in A_{rn}} c_a^t\left(f_a^t\right) d_{ra}^{d\,t} \quad \forall\ r \in R, n \in N, t \in T \tag{14-55}$$

$$\left[b_{rn}^t - t\ \Delta t\right] d_{ra}^{d\ t} \leq 0 \quad \forall\ r \in R, n \in N, d\ \in T, t \in T, a \in A_n \tag{14-56}$$

$$\left[b_{rn}^t - (t-1)\ \Delta t\right] d_{ra}^{d\ t} \geq 0 \quad \forall\ r \in R, n \in N, d \in T, t \in T, a \in A_n \tag{14-57}$$

14.4 SUO-based Approaches

14.4.1 *SUE NA (Fisk 1980)*

Assumptions:

i) one mode;

ii) separable cost functions (12-59);

iii) non-negative cost functions (12-63);

iv) increasing cost functions (12-64);

v) D_{ij} fixed and exogenous;

vi) route costs are random variables consisting of an observable or structural component and an unobservable or stochastic component whose expected value is zero (13-10), (13-11).

Optimization problem:

$$\underset{\{h_r,\, f_a\}}{MIN} \quad \frac{1}{q} \sum_{i \in I} \sum_{j \in J} \sum_{r \in R_{ij}} h_r \ln h_r + \sum_{a \in A} \int_0^{f_a} c_a(x)\, dx \tag{14-58}$$

subject to:

$$\sum_{r\in R_{ij}} D_{ijr} = D_{ij} \ \forall\ i, j \qquad \textbf{(14-59)}$$

$$\sum_{i\in I}\sum_{j\in J}\sum_{r\in R_{ij}} d_{ar} h_r = f_a \quad \forall a \in A \qquad \textbf{(14-60)}$$

$$h_r \geq 0 \quad \forall\ r \in R_{ij} \quad \forall\ i \in I \quad \forall j \in J \qquad \textbf{(14-61)}$$

14.4.2 *Super- and hyper-networks (Sheffi and Daganzo 1980)*

Assumptions:

i) $$\overline{G} = \left[\overline{N}, \overline{A}\right] \qquad \textbf{(14-62)}$$

ii) $$\overline{N} = N \cup V \qquad \textbf{(14-63)}$$

iii) $$\overline{A} = A \cup E \qquad \textbf{(14-64)}$$

iv) c_a fixed and exogenous $\forall a \in E$;

v) separable cost functions(12-59) $\forall\ a \in A$;

vi) non-negative cost functions (12-63) $\forall\ a \in A$;

vii) increasing cost function (12-64) $\forall\ a \in A$;

Hypernetwork equilibrium conditions:

$$D''_{lm} = \sum_{i\in I}\sum_{j\in J} D_{ij} p^{ij}_{lm} \quad \forall\ l \in N', m \in N'' \qquad \textbf{(14-65)}$$

$$\sum_{r\in R_{lm}} h_r = D''_{lm} \quad \forall\ l \in N', m \in N'', i \in I, j \in J \qquad \textbf{(14-66)}$$

$$C_r - C_{lm*} \geq 0 \quad \forall\ r \in R_{lm}, \forall l \in N', m \in N'' \qquad \textbf{(14-67)}$$

$$\left(C_r - C_{lm^*}\right) h_r = 0 \quad \forall\ r \in R_{lm}, l \in N', m \in N'' \qquad \textbf{(14-68)}$$

$$h_r \geq 0 \quad \forall\ r \in R \qquad \textbf{(14-69)}$$

where:

$p^{ij}_{lm} =$	$\Pr ob\left[\overline{P}^{ij}_{lm} \leq \overline{P}^{ij}_{no} \quad \forall\, n \in N'_i, o \in N''_j\right]$	**(14-70)**
$\overline{P}^{ij}_{lm} =$	$P^{ij}_l + C_{lm^*} + P^{ij}_m$	**(14-71)**
$P^{ij}_l =$	stochastic cost on entrance arc $(n_i, n_l) \in E$	**(14-72)**
P^{ij}_m	stochastic cost on egress arc $\left(n_m, n_j\right) \in E$	**(14-73)**
$C_{lm*} =$	minimum path cost between basic network entrance/egress pair *l,m*	**(14-74)**

14.5 Combined UO-S/SUO Approaches

14.5.1 *Combined TG/TD/MS/NA - Trip Consumer Approach (Oppenheim 1995)*

Assumptions:

i) one or more modes;

ii) non-negative cost functions (12-63);

ii) increasing cost functions (12-64);

iv) separable cost functions (12-59) (although Oppenheim (1995) discusses a two-mode non-separable cost function version of the model).

Consumer utility maximization problem (Varian 1992):

$$\underset{\{\mathbf{y}\}}{MAX} \quad U(\mathbf{y}) \qquad \textbf{(14-75)}$$

subject to:

$$\mathbf{p}\;\mathbf{y} = b \qquad \textbf{(14-76)}$$

$$\mathbf{y} \in \mathbf{Y} \qquad \textbf{(14-77)}$$

where U(**y**) is the utility of choice **y**, **y** is an attribute vector, **p** is a vector of prices associated with each attribute, b is a budget constraint and **Y** is the feasible solution space.

Indirect and expected utilities:

TG	Indirect utility:	$\tilde{U}_{i\|l} = b_t \, (V_i + \tilde{W}_{i\|l+1})$	**(14-78)**
	Expected utility	$\tilde{W}_{i\|l} = \dfrac{1}{b_t} \ln\left[1 + \exp\left(b_t\left(V_i + \tilde{W}_{i\|l+1}\right)\right)\right]$	**(14-79)**
TD	Indirect utility	$\tilde{U}_{ij\|l} = b_d \, (V_{ij} + \tilde{W}_{ij\|l+1})$	**(14-80)**
	Expected utility	$\tilde{W}_{ij\|l} = \dfrac{1}{b_d} \ln\left[\sum_j \exp\left(b_d\left(V_{ij} + \tilde{W}_{ij\|l+1}\right)\right)\right]$	**(14-81)**
MS	Indirect utility	$\tilde{U}_{ijm\|l} = b_m \, (V_{ijm} + \tilde{W}_{ijm\|l+1})$	**(14-82)**
	Expected utility	$\tilde{W}_{ijm\|l} = \dfrac{1}{b_m} \ln\left[\sum_m \exp\left(b_m\left(V_{ijm} + \tilde{W}_{ijm\|l+1}\right)\right)\right]$	**(14-83)**
NA (UO-S)	Indirect utility	$g_{ijmr} = - t \, t_{ijmr} - c_{ijmr}$	**(14-84)**
	Expected utility	$g^*_{ijmr} = \mathrm{MIN}_{\{k\}} \, g_{ijmk}$	**(14-85)**
NA (SUO)	Indirect utility	$\tilde{U}_{ijmr\|l} = b_r \, g_{ijmr}$	**(14-86)**
	Expected utility	$\tilde{W}_{ijmr\|l} = \dfrac{1}{b_r} \ln\left[\sum_r \exp\left(b_r \, g_{ijmr}\right)\right]$	**(14-87)**

Gormon-form utility structure (Varian 1992):

$$U\left(\mathbf{p}, \mathbf{z_n}, b_n\right) = f\left(\mathbf{p}, \mathbf{z_n}\right) + b_n \, g(\mathbf{p}) \qquad \textbf{(14-88)}$$

where:

$\mathbf{p}$ = vector of observed prices or costs;

$\mathbf{z_h}$= vector of observed attributes for individual *n*;

b_n= budget of individual *n*.

Direct utilities (objective function components):

TG:	$\frac{1}{b_t^l}\sum_{i\in I}\left(D_i \ln D_i + D_{i0}\ln D_{i0}\right) - \sum_{i\in I} U_i D_i$	**(14-89)**
TD:	$\frac{1}{b_d^l}\sum_{i\in I}\sum_{j\in J} D_{ij}\ln D_{ij} - \sum_{i\in I}\sum_{j\in J} U_{ij} D_{ij}$	**(14-90)**
MS:	$\frac{1}{b_k^l}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} D_{ij}^k \ln D_{ij}^k - \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} U_{ij}^k D_{ij}^k$	**(14-91)**
NA-D	$t\sum_{k\in K}\sum_{a\in A}\int_0^{f_a^k} c_a^k(x)dx$	**(14-92)**
NA-S	$\frac{1}{b_r^l}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\sum_{r\in R_{ij}} h_r^k \ln h_r^k + t\sum_{k\in K}\sum_{a\in A}\int_0^{f_a^k} c_a^k(x)dx$	**(14-93)**

Travel demand constraints:

TGC:	$D_i + D_{i0} = B_i \quad \forall\, i \in I$	**(14-94)**
TDC:	$\sum_{j\in J} D_{ij} = D_i \quad \forall\, i\in I$	**(14-95)**
MSC:	$\sum_{k\in K} D_{ij}^k = D_{ij}$	**(14-96)**
NAC:	$\sum_{r\in R_{ij}} h_r^k = D_{ij}^k \quad \forall\, i\in I, j\in J, k\in K$	**(14-97)**

Non-negativity constraints:

$D_{i0} \geq 0 \quad \forall\, i\in I$	**(14-98)**
$D_{ij} \geq 0 \quad \forall\, i\in I, j\in J$	**(14-99)**
$D_{ij}^k \geq 0 \quad \forall\, i\in I, j\in J, k\in K$	**(14-100)**
$h_r^k \geq 0 \quad \forall\, r\in R, k\in K$	**(14-101)**

Parameter restrictions:

$$b_k^l = \begin{cases} \dfrac{b_k^l b_k^{l+1}}{b_k^{l+1} - b_k^l} & , l = 1, \ldots, L \\ b_k & , 1 = L \end{cases} \tag{14-102}$$

Equation (14-102) implies:

$$b_k^1 \leq \ldots \leq b_k^L \tag{14-103}$$